AUTHENTIC ASSESSMENT AND ITS IMPACT ON SEAFARER STUDENTS' ACADEMIC ACHIEVEMENT: A COMPARATIVE ANALYSIS WITH TRADITIONAL ASSESSMENT

Samrat Ghosh
Master of International Business
Master of Business Administration (Maritime and Logistics Management)
Master Mariner
Bachelor's in Marine Science
Certificate IV in Training and Assessment

Supervisors:
Associate Professor Benjamin Brooks
Professor Dev Ranmuthugala
Professor Marcus Bowles

Submitted in partial fulfilment of the requirements for the Degree of Doctor of Philosophy
University of Tasmania, Australian Maritime College
February 2020

DECLARATION OF ORIGINALITY

This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of the thesis, nor does the thesis does not contain any material that infringes copyright.

_____ Date: 02/02/2020

Samrat Ghosh

STATEMENT OF ETHICAL CONDUCT

The research associated with this thesis abides by the *Australian Code for the Responsible Conduct in Research* (2018) and the *National Statement on Ethical Conduct in Human Research* [2007 (Updated 2018)].

_____ Date: 02/02/2020

Samrat Ghosh

STATEMENT REGARDING PUBLISHED WORK AND AUTHORITY OF ACCESS

The publishers of the papers comprising Chapter 7 hold the copyright for that content, and access to the material should be sought from the respective journals. The remaining non-published content of the thesis may be made available for loan, limited copying and communication in accordance with the *Copyright Act* (1968).

_____ Date: 02/02/2020

Samrat Ghosh

STATEMENT OF CO-AUTHORSHIP

The following people and institutions contributed to the publication of work undertaken as part of the thesis:

Candidate -- Samrat Ghosh, Australian Maritime College, University of Tasmania, Australia.

Author 1 -- Benjamin Brooks, Australian Maritime College, University of Tasmania, Australia.

Author 2 -- Dev Ranmuthugala, Defence Science and Technology Group, Australia.

Author 3 -- Marcus Bowles, Deakin University, Australia.

Contribution of work by co-authors for each paper:

Paper 1: Located in Chapter 7 (Section 7.1)
Ghosh, S., Bowles, M., Ranmuthugala, D. and Brooks, B. (2014), "Reviewing seafarer assessment methods to determine the need for authentic assessment", *Australian Journal of Maritime & Ocean Affairs*, Vol. 6, No. 1, pp. 49-63.

Author contributions:
Main author and primarily responsible for the conception, planning and execution of the review of past literature: Candidate.
Contributed to the idea as well as its formalisation and refinement: Author 1, Author 2, Author 3.

Paper 2: Located in Chapter 7 (Section 7.2)
Ghosh, S., Bowles, M., Ranmuthugala, D. and Brooks, B. (2014), "On a lookout beyond STCW: Seeking standards and context for the authentic assessment of seafarers", in Ranmuthugala, D. and Lewarn, B. (Eds.), *IAMU AGA 15 Looking Ahead: Innovation in Maritime Education, Training and Research, Launceston, Australia, 27-30 October*.

Author contributions:
Main author and primarily responsible for the conception, planning and execution of the review of past literature: Candidate.
Contributed to the idea as well as its formalisation and refinement: Author 1, Author 2, Author 3.

Paper 3: Located in Chapter 7 (Section 7.3)
Ghosh, S., Bowles, M., Ranmuthugala, D. and Brooks, B. (2015), "Using authentic assessment to enhance seafarer student engagement and their ability to transfer learning", in Svilicic, B. and Pritchard, B. (Eds.), *IAMU AGA 2015 International Association of Maritime Universities 16th IAMU Annual General Assembly, Opatija, Croatia, 7-10 October*.

Author contributions:
Main author and primarily responsible for the conception, planning and execution of the review of past literature: Candidate.
Contributed to the idea as well as its formalisation and refinement: Author 1, Author 2, Author 3.

Paper 4: Located in Chapter 7 (Section 7.4)
Ghosh, S., Bowles, M., Ranmuthugala, D. and Brooks, B. (2016), "Authentic assessment in seafarer education: Using literature review to investigate its' validity and reliability through rubrics", *WMU Journal of Maritime Affairs*, Vol. 15, No. 2, pp. 317-330.

Author contributions:
Main author and primarily responsible for the conception, planning and execution of the work: Candidate.
Executed the literature review: Candidate.
Contributed to the idea as well as its formalisation and refinement: Author 1, Author 2, Author 3.

Paper 5: Located in Chapter 7 (Section 7.5)
Ghosh, S., Bowles, M., Ranmuthugala, D. and Brooks, B. (2017), "Improving the validity and reliability of authentic assessment in seafarer education and training: A conceptual and practical framework to enhance resulting assessment outcomes", *WMU Journal of Maritime Affairs*, Vol. 16, No. 3, pp. 455-472.

Author contributions:
Main author and primarily responsible for the conception, planning and execution of the work: Candidate.
Executed the literature review and the formulation of the conceptual framework: Candidate.
Contributed to the idea as well as its formalisation and refinement: Author 1, Author 2, Author 3.

Paper 6: Located in Chapter 7 (Section 7.6)
Ghosh, S., Brooks, B., Ranmuthugala, D. and Bowles, M. (2019), "Authentic versus traditional assessment: An empirical study investigating the difference in seafarer students' academic achievement", accepted for publishing in Journal of Navigation.

Author contributions:
Main author and primarily responsible for the conception, planning and execution of the work: Candidate.
Conceived and designed the experiment: Candidate, Author 1.
Analysed the quantitative data: Candidate.
Contributed to the idea as well as its formalisation and refinement: Author 1, Author 2, Author 3.

Paper 7: Located in Chapter 7 (Section 7.7)
Ghosh, S., Brooks, B., Ranmuthugala, D. and Bowles, M. (2019), "Investigating the correlation between students' perceptions of authenticity in assessment and their academic achievement in the associated assessment tasks", under review in Journal of Navigation.

Main author and primarily responsible for the conception, planning and execution of the work: Candidate.
Conceived and designed the experiment: Candidate, Author 1.
Analysed the quantitative data including the correlational analysis: Candidate.
Contributed to the idea as well as its formalisation and refinement: Author 1, Author 2, Author 3.

We, the undersigned, endorse the above-stated contribution of work undertaken for each of the published (or accepted or submitted) peer-reviewed manuscripts contributing to this thesis:

Signed: _____ _____

Samrat Ghosh	Benjamin Brooks	Michael van Balen
Candidate	Primary Supervisor	Head of School
Australian Maritime College	Australian Maritime College	Australian Maritime College
University of Tasmania	University of Tasmania	University of Tasmania
Date: 02/02/2020	02/02/2020	03/02/2020

ACKNOWLEDGEMENTS

I would like to begin by thanking the unknown force that I know as the almighty God, who gave me the strength and the courage to not only begin but also to complete this PhD. I also believe that the blessing of my parents made, what seemed like an impossible dream, come true.

My sincere gratitude to the supervisory team that comprised of Benjamin Brooks, Dev Ranmuthugala and Marcus Bowles. I am indebted to them for their invaluable supervision, patience, and inputs that instilled in me the confidence and belief to complete this research project. They inspired me to go beyond my capability and aspirations. My supervisors should also be credited for the improvement in my academic writing skills and my evolution as a researcher and academic. I especially thank Marcus who not only encouraged me to begin but was the primary supervisor for most of my journey. Apart from being a supervisor, Marcus was a friend, philosopher, and guide who helped me navigate through the fascinating world of research, making it interesting and inspirational.

This research could not have been completed without the support of many people within the Australian Maritime College and the University of Tasmania. I am grateful to Mark Symes, Barrie Lewarn, Cecile L'Hermitte, Ian Bollard, Doug Colbeck, and Katherine Shaw. A special note of thanks to Hossein Enshaei for being very helpful with his valuable advice and guidance as the graduate research coordinator, a colleague and fellow academic. My heartfelt thanks to Michael Douglas and Alpha Isaako in providing me with the required support in the implementation of the authentic assessment. In addition to the university staff, special thanks to all the survey respondents who took part in this research.

I would also like to acknowledge my friends and family who kept me sane whenever I felt lost and confused. Finally, this journey would not have been possible without the constant support, love, and strength of my beautiful wife Aparna and son Aditya. No words can express the gratitude I feel towards them for being patient with me and tolerating my mood swings and loving me selflessly. I thank Aparna for being by my side since the beginning of the journey and listening patiently to the various versions of my writing and research thoughts. I express my appreciation to both Aparna and Aditya who made numerous sacrifices to help me fulfil my dream. I dedicate this thesis to them.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

The following abbreviations are used in this thesis:

α	Cronbach's Alpha
AA	Authentic Assessment Scores (Combined)
AA_1	Authentic Assessment Scores (First Task)
AA_2	Authentic Assessment Scores (Second Task)
AAFSET	Authentic Assessment Framework for Seafarer Education and Training
AGCS	Allianz Global and Corporate Specialty
AMC	Australian Maritime College
AMSA	Australian Maritime Safety Authority
CBT	Competency-based Training
CoC	Certificate of Competence
df	Degree of Freedom
GPS	Global Positioning System
FSC	Flag State Control
H_1	Hypothesis for Research Question 1
H_2	Hypothesis for Research Question 2
ILOs	Intended Learning Outcomes
IMDG	International Maritime Dangerous Goods
IMO	International Maritime Organization
IMSBC	International Maritime Solid Bulk Cargoes
ISM	International Safety Management
MET	Maritime Education and Training
MCQ	Multiple-choice questions
NSSE	National Survey of Student Engagement
OBE	Outcomes-based Education
PSC	Port State Control
RQ	Research Question
SD	Standard Deviation
SET	Seafarer Education and Training
SPSS	Statistical Package for the Social Sciences
STCW	Standards of Training, Certification, and Watchkeeping
TA	Traditional Assessment Scores (Combined)
TA_1	Traditional Assessment Scores (First Task)
TA_2	Traditional Assessment Scores (Second Task)
UTAS	University of Tasmania
VET	Vocational Education and Training

LIST AND STATUS OF PUBLICATIONS

This research includes seven papers. Five papers have been published and two are under review. These papers have been reproduced in Chapter 7.

Paper	Nature	Status	Title	Publication channel	Publication details	Full-length, double-blind review
I	Conceptual paper	Published (2014)	Reviewing seafarer assessment methods to determine the need for authentic assessment	*Australian Journal of Maritime & Ocean Affairs*	Vol. 6, No. 1, pp. 49-63	Yes
II	Conceptual paper	Published (2014)	On a lookout beyond STCW: Seeking standards and context for the authentic assessment of seafarers	IAMU AGA 15 Looking Ahead: Innovation in Maritime Education, Training and Research	ISBN number: 978-0-9806391-4-8 Editors: Ranmuthugala, D. and Lewarn, B	Yes
III	Conceptual paper	Published (2015)	Using authentic assessment to enhance seafarer student engagement and their ability to transfer learning	IAMU AGA 2015 International Association of Maritime Universities 16th IAMU Annual General Assembly	ISBN number: 978-953-165-116-5 Editors: Svilicic, B. and Pritchard, B.	Yes
IV	Conceptual/ Research paper	Published (2016)	Authentic assessment in seafarer education: Using literature review to investigate its' validity and reliability through rubrics	*WMU Journal of Maritime Affairs*	Vol. 15, No. 2, pp. 317-330	Yes
V	Conceptual/ Research paper	Published (2017)	Improving the validity and reliability of authentic assessment in seafarer education and training: A conceptual and practical framework to enhance resulting assessment outcomes	*WMU Journal of Maritime Affairs*	Vol. 16, No. 3, pp. 455-472	Yes
VI	Research paper	Accepted for publishing	Authentic vs traditional assessment: an empirical study investigating the difference in seafarer students' academic	*Journal of Navigation*	Not yet applicable	Yes
VII	Research paper	Under review	Investigating the correlation between seafarer students' perception of authenticity in assessment and their academic achievement in the associated assessment tasks	*Submitted and under review in Journal of Navigation*	Not yet applicable	Yes

ABSTRACT

The Standards of Training, Certification and Watchkeeping (STCW) Convention set global, minimum standards of competence for seafarers in 1978 by introducing the STCW Code (STCW'78). Through amendments in 1995 and 2010, the convention intended to improve the knowledge-based training mandate established in STCW'78 by making it outcome-based. With STCW moving seafarer training towards outcomes-based education (OBE), emphasis shifted to assessment practices that will allow seafarer students to demonstrate their ability to perform workplace tasks at standards described in the revised and current STCW'95 Code (including 2010 amendments). Seafarer education and training (SET) institutes working under the directives of the national maritime regulator [e.g. Australian Maritime Safety Authority (AMSA) in Australia] are responsible for ensuring that the adopted assessment methods, as promoted and recommended by the STCW Code, not only assure attainment of STCW outcomes but produce competent graduates that meet the expectations of the core stakeholders such as regulators and employers.

However, a critical review of specific excerpts from the STCW Code was used in this research study to show that the Code largely fails to provide a 'standard' that can assure assessment of seafarers to one of the most critical outcomes: the performance expected at a level of work in the industry. Moreover, a review of past research in the area of seafarer education conducted for this study, showed that the traditional assessments that required seafarer students to focus on rote learning and construction of responses devoid of real-world contexts (e.g. oral examinations, written assignments, and multiple-choice questions) disengaged them from learning. Memorising information is a lower-order cognitive ability, failure in which led to errors in assessment tasks resulting in low academic achievement for students. Hence, this research proposed that authentic assessments, requiring students to construct responses based on the assimilation, integration, and critical analysis of information presented in real-world contexts will result in higher academic achievement.

Using the characteristics recommended by the commonly cited authors, this study redefined the concept of authentic assessment which established the theoretical framework for this study. However, authentic assessment can capture essential aspects of workplace tasks and result in consistency of student performance in different contexts only if they are valid and reliable. Rubrics as assessment tools are known to increase validity and reliability of assessments, but it can do so only if different aspects of its own validity and reliability have been addressed. An extensive literature review in the area of authentic assessment, conducted as part of this research study, uncovered an absence of academic investigation and empirical study on the different aspects of validity and reliability of authentic assessment, when implemented with as well as without the use of assessment rubrics. In this regard, a conceptual and practical authentic assessment framework for seafarer education and training (AAFSET) that promotes a holistic approach and provides greater assurances

of validity and reliability throughout all stages of assessment within seafarer programs was developed in this research study.

The findings of the literature review also revealed that there existed an even greater absence of global research on authentic assessment in the area of seafarer training. Hence, an empirical contribution of this study was through the investigation of the difference in seafarer students' academic achievement (measured through scores obtained in assessment) in authentic assessment as compared with traditional assessment. Two separate and independent student groups as the 'control' and 'treatment' group were used for a selected unit of learning delivered at the Australian Maritime College (AMC) within the Bachelor of Nautical Science degree program. Since, some past researchers defined traditional assessments as a single-occasion assessment implemented at the end of the learning period, this project implemented the assessment in a summative format as opposed to authentic assessments implemented formatively. Analysis of student scores revealed that the authentically assessed students were guided towards significantly higher academic achievement.

A further investigation using the students undergoing authentic assessment, included measuring their perceptions of authenticity for factors of assessment (task, criteria, etc.) and correlating to their scores in the associated task. Stage 1 focused on deriving the factors conceptually using the definition of authentic assessment, based on which a perception survey questionnaire was designed. Following the collection of student responses through the survey, a correlational analysis was conducted between student perceptions and their scores. Stage 2 extracted new factors through a factor analysis. Using the survey data (but for the new factors), an additional correlational analysis was conducted to confirm findings. Both stages of investigation found that the factor of transparency of criteria was a significant predictor of the students' academic achievement.

Future research will investigate seafarer students' perceptions through the use of qualitative methodologies (e.g. interviews and focus groups) to gain an in-depth knowledge of other factors that may enhance authenticity of assessments. Project findings identified vital challenges for the implementation of authentic assessment and uncovered significant factors of assessment which, if included, in the design of the assessment will guide authentically assessed students towards higher academic achievement.

1. INTRODUCTION

Chapter 1 establishes the research background. It introduces and expands on the main concepts of this research project by providing definitions of the key terms and stakeholders involved. This chapter also defines the research problem and in doing so identifies the research gaps and formulates a set of research questions to guide this doctoral thesis. In terms of publications included in this thesis, the research problem was established in Paper I and Paper II, the research gaps were identified in Paper IV and Paper V, and the research questions were formulated in Paper VI and Paper VII.

1.1. Background

1.1.1. The basis of assessment methods currently used in seafarer education

Seafarers are entrusted with ships and cargo worth billions of dollars and the lives of passengers and their colleagues which are priceless. Accidents in the seafaring industry can have catastrophic effects leading to loss of lives and major environmental damage. Analysis of accidents have often revealed that the lack of sufficient competence to operate a ship and its systems have often contributed to such accidents (Pecota & Buckley 2009). For example, the analysis of worldwide shipping losses (of ships over 100 gross tons) in 2014 by Allianz Global Corporate & Specialty (AGCS 2015) revealed that while safety is gradually improving, the lack of sufficient competence to operate a ship and its systems is a major factor in either causing or contributing to maritime accidents and safety breaches. The seafaring industry will always be at risk of major operational errors if the competence levels of seafarer students are not accurately and adequately assessed. The assessment of seafarer students' ability to perform tasks at workplace standards should be conducted before issuing them the licence (Certificate of Competence or CoC) required to operate ships at different levels of responsibility.

The International Maritime Organisation (IMO) introduced the Standards of Training, Certification and Watchkeeping (STCW) Convention in 1978 (referred to as STCW'78) with one of its primary objectives aimed at reducing human error due to inadequate competence and training. In a significant step by the IMO, the human error problem was viewed through the 'person approach' (Reason 2000), making the failure of actions to be primarily arising from informational problems (e.g. inadequate knowledge, unfamiliarity with situation, and lack of competence) (Reason 1995). However, accidents occur due to the outcome of complex interactions and coincidences of systems and individuals and cannot be explained in simplistic cause-effect terms where individual operators can be solely blamed for its cause (Reason, Hollnagel & Paries 2006).

Seafarers on board ships constantly perform challenging operations that require them to interact with systems and technology in complex scenarios to achieve their desired goals. This has much potential for human error (Ernstsen & Nazir 2018). For example, navigating deep-draft ships in relatively shallow waters and reduced visibility with the improper use of the radar equipment and insufficient use of personnel for lookout/steering may lead to collision or grounding. In the seafaring industry, technical errors are less prone than human errors, which means the limited variety of technical or process errors allow training and assessment to be proactive in skilling seafarers to deal with critical scenarios (Ernstsen & Nazir 2018). This is because individual operators can reduce accidents at the workplace and reduce human error (Nazir et al. 2015). This can be achieved if training for competence development includes simulation of scenarios resembling the workplace, thereby providing situational cues enabling students to practice dealing with unanticipated situations (Kluge 2008; Kluge, Nazir & Manca 2014; Flin, O'Connor & Crichton 2008). The training should address day-to-day shipboard operations, risk assessment procedures as well as contingency planning and emergency preparedness (Kim & Nazir 2016; Kim, Nazir & Øvergård 2016).

The IMO intended to reduce human error due to inadequate competence by establishing global, minimum standards of competence through the STCW Code. The code provided guidelines to seafarer education and training (SET) institutes [or maritime education and training (MET) institutes] for assessing seafarers via a standardised system of competence assessment. Based on the students' performance in assessments, the latter can be granted approval to apply for a final assessment by the national maritime regulators (e.g. the Australian Maritime Safety Authority or AMSA in Australia) in order to obtain their CoC. The CoC opens job opportunities for seafarer students within seamless career pathways that, based on competence, can take them from entry level to ship's captain and beyond. National maritime regulatory bodies use the STCW approved CoC as a basis for regulation of the workforce on ships registered under their flags. Employers consider the CoC to be evidence of a seafarer's competence to perform at the workplace and a basis for them to recruit, reward, and train their employees. Thus, the assessment process leading to the issuance of the CoC, becomes a key component to fulfil the expectations associated with it for seafarer students, employers, SET institutes, and maritime regulatory bodies.

In education, assessment can be defined as "a systematic collection, review, and use of" (Walvoord 2004, p. 2) data that can be conducted at various stages of the learning cycle to acquire feedback about: a student's progress and achievements, the effectiveness of teaching and instruction, and the attainment of course outcomes (University of Tasmania [UTAS] 2011), while fulfilling the overall goal of improving student learning (Palomba & Banta 1999). In outcomes-based education (OBE) such as vocational education and training (VET) or competency-based training (CBT), assessments also provide feedback

about the attainment by students of minimum standards that are essentially required for the workplace (Brady 1997). Standards in such cases become the outcomes (Burke 2011) or more correctly 'learning outcomes' establishing what the students should be able to demonstrate at the end of the learning period (Driscoll & Wood 2007). Students direct their learning efforts towards 'outcome' attainment and assessors are guided on what they are supposed to measure via assessments. The evidence produced from the assessments can be used by educators to not only improve teaching practices by identifying learning needs, but also to meet accountability requirements by providing feedback to stakeholders on the learners' progress towards the achievement of standards (Brindley 1998).

In the context of professional education and training, such as seafarer education, assessment also provides feedback about the achievement by students of professional standards (as provided by the STCW Code) that are essential for the workplace. Attainment of such standards provide evidence of the ability to combine knowledge, skills, values, and attitudes into behaviours (Aranda & Yates 2009) required to perform in the real world. This ability also defines the student's professional competence (Rychen 2004). Professional competence requires essential cognitive abilities of recalling information (knowledge) and applying it (skills) based on analytical and critical thinking (Nusche 2008). Underlying are the values and attitudes that are non-cognitive skills that shape the principles of thought and prompt responses based on the reflection of those thoughts (Moore & Asay 2013). Seafarers can acquire professional competence in workplace settings on board ships and in SET institutes under academic guidance as the training structure of a seafarer is alternatively divided into college and sea-based training.

The IMO is the over-arching regulatory body that together with national maritime regulators, enforces the requirements of the STCW on maritime nations (or states) that have ratified the convention. The national regulator is usually known as the 'Flag State' (also known as Flag State Control or FSC), which can be an administration or the government of the state under which ships can be registered, (e.g. AMSA in Australia). The FSC becomes the 'Port State Control' (PSC) when ships of other registry call at their ports. The FSC and PSC of nations that have ratified STCW ensure compliance through inspections and surveys for all ships irrespective of registry when they visit ports of signatory nations. Countries fully complying with the provisions of STCW and its education practices are listed as the 'white list' of nations in the maritime industry. The IMO has authorised the national maritime regulators to grant approval to the SET institutes for conducting STCW approved courses if, after a thorough inspection of their facilities and processes, they are found to be in full compliance with the provisions of the STCW code. Table 1.1 describes the roles of the key stakeholders in the implementation and adoption of the STCW Code.

Table 1.1: The role of the key seafaring stakeholders in the assessment process.

Key seafaring stakeholders	Roles in the assessment process
International Maritime Organization (IMO)	• Overarching regulatory body. • Introduced the STCW Code to regulate the standards of competence for seafarer students.
Maritime Regulators: Flag State Control (FSC); and Port State Control (PSC)	• Authorities responsible for the implementation of the STCW Code on the maritime nations. • Inspect seafarer education training institutes (SET) to ensure compliance with the STCW Code in training and assessment. • Conduct final assessment of seafarer students towards granting the certificate of competence (CoC).
Seafarer Education and Training (SET) Institutes	• Conduct training and assessment of seafarer students based on the STCW Code. • Based on performance in assessments, grant approval to seafarer students to apply for a final assessment by maritime regulators towards granting the CoC.
Seafarer Students	• Undergo training and assessment at SET towards obtaining CoC to serve at different levels of responsibility on ships.
Seafarer Employers	• Recruits, rewards, and trains the seafarer students on the basis of the CoC.

The STCW Code applies to seafarers who are working or intending to work on commercial vessels on domestic (coastal) or international voyages but not to those serving on naval vessels, government-owned or operated vessels engaged in non-commercial service, fishing vessels, pleasure yachts not engaged in trade and wooden ships of primitive build (STCW 2011). Over the years, the STCW has been updated with various amendments in 1997, 1998, 2004, 2006, and in 2010 with the Manila amendments.

Prior to the introduction of STCW'78, individual countries established their own standards. However, the STCW'78 did not prove to be as effective as expected due to criticisms from stakeholders that vague and unclear standards were left to individual interpretation by maritime nations (IMO 2013b), which posed the risk of variation in the standards of competence development among international seafarers. Moreover, the STCW'78 was essentially knowledge-based comprising a syllabus for a quantifying examination instead of focusing on skills and abilities necessary to perform workplace tasks (Morrison 1997). The IMO revised the STCW Code through the 1995 amendments (known as STCW'95) intending to fundamentally improve the training mandate by making it outcome-based. This would require seafarers to demonstrate their competence in the tasks outlined in the STCW Code rather than just show they had acquired knowledge (as stipulated in STCW'78). Currently, the revised STCW Code is referred to as STCW'95 including the 2010 Manila Amendments aimed at bringing the STCW convention and Code up to date with industry developments. Several areas of the Manila amendments that anticipated and addressed needs to emerge in the foreseeable

future, e.g. teaching and assessment of essential soft and underlying competencies of leadership and team working. Figure 1.1 outlines the structure of the STCW'95 Code (after its last revision in 2010). Table 1.2 describes the standards of competence and assessment as laid out in STCW'95 (after its last revision in 2010).

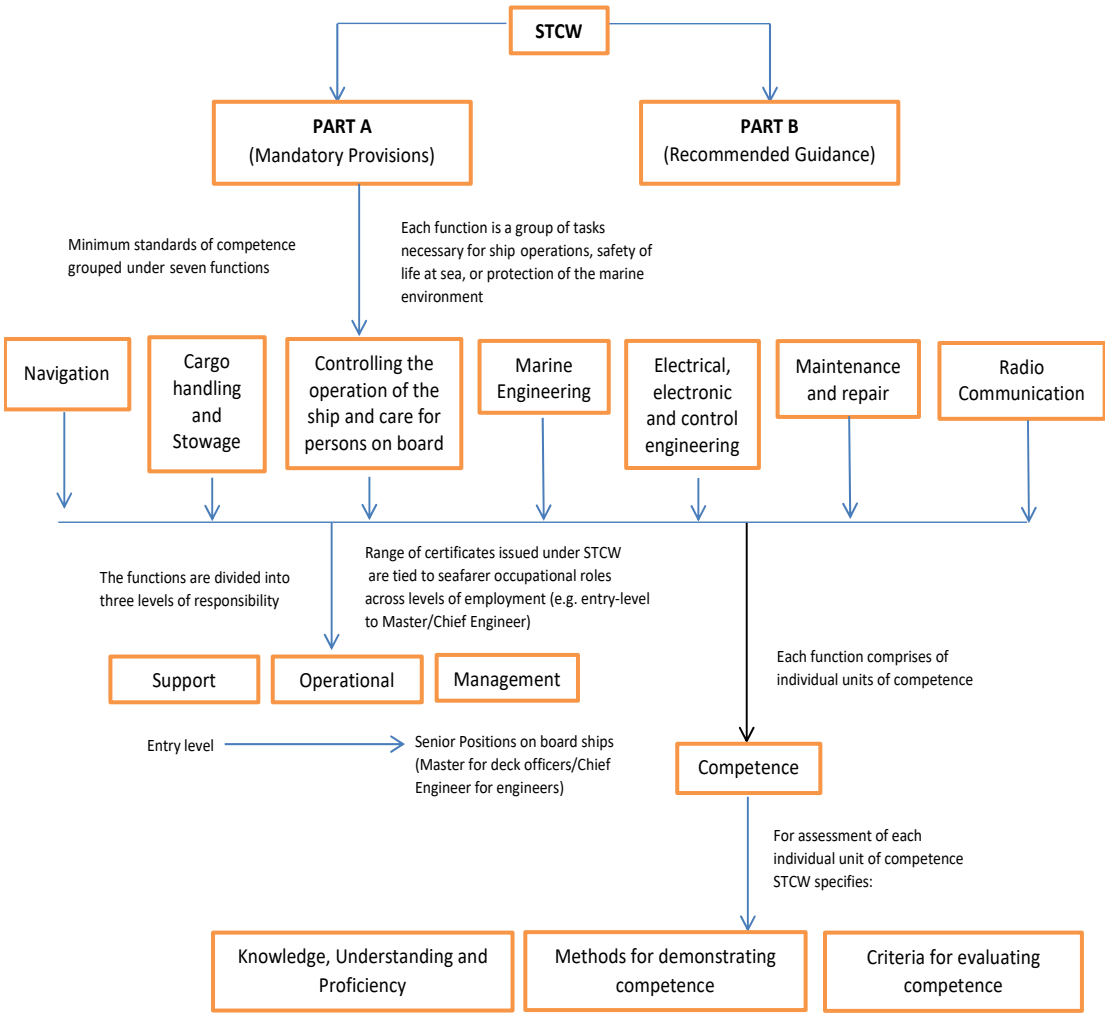


Figure 1.1: Structure of the STCW'95 Code (including the 2010 revision).

Table 1.2: Example of the STCW Code defines the minimum standards for competence assessment for the 'Carriage of dangerous goods' (IMO 2011).

Competence	**Knowledge, understanding and proficiency**	**Methods for demonstrating competence**	**Criteria for evaluating competence**
Carriage of dangerous goods	International regulations, standards, codes and recommendations on the carriage of dangerous cargoes, including the International Maritime Dangerous Goods (IMDG) Code and the International Maritime Solid Bulk Cargoes (IMSBC) Code Carriage of dangerous, hazardous and harmful cargoes; precautions during loading and unloading and care during the voyage	Examination and assessment of evidence obtained from one or more of the following: .1 approved in-service experience .2 approved simulator training, where appropriate .3 approved specialist training	Planned distribution of cargo is based on reliable information and is in accordance with established guidelines and legislative requirements Information on dangers, hazards and special requirements is recorded in a format suitable for easy reference in the event of an incident

Typically working under the oversight of the maritime regulator, it is the responsibility of the SET institutes to assess and collect evidence of students' ability to apply knowledge learned in the classroom and during seatime (service on board a ship) to practical skills at appropriate levels of responsibility (e.g. support, operational and management) in accordance with the STCW Code.

1.2. Problem definition

Although, much of the SET institutes uses simulators and practical exercises for training and assessment in selected units of the STCW Code (Sellberg 2017), the use of decontextualised traditional assessment methods for most of the units of competence listed in the STCW Code cannot be ignored. Past research (Maringa 2015; Emad & Roth 2007; AMC 2011) has demonstrated that seafarer students tend to disengage with traditional assessments (e.g. multiple-choice questions (MCQ), oral examinations, and written assignments) presented devoid of real-world contexts and that focus only on their ability to recall and regurgitate the body of knowledge taught in the classroom. Disengaged students opt for surface-learning approaches (Maltby & Mackie 2009) relying on rote learning instead of assimilating and analysing information critically towards preparation for such assessment tasks. For example, one of the ways a seafarer is

certified as competent to work onboard commercial ships is through an assessment based on memorised answers in an oral examination. Traditional assessments like oral examinations required students to construct responses devoid of real-world contexts relying solely on the students' ability to visualise work-based scenarios. However, the ability to memorise is a lower-level cognition, and memory lapses may lead to unintentional skill and knowledge-based errors (Wiggins 1990) leading to poor academic achievement. Although one may argue that traditional assessments like MCQ and oral examinations can also be authentic in particular contexts, Mueller (2006) suggests that they are on the lower end of the continuum of authenticity when they focus on the attributes of recall and recognition. Traditional assessment methods may be effective in assessing knowledge-based components of a task but, they are somewhat decontextualised in nature, and it is difficult to provide students with a real-world context for skills and knowledge application (Boud & Falchikov 2006; Kearney 2012).

1.2.1. Lack of student engagement with traditional assessments

An ethnographic case study involving 16 seafarer students in a Canadian maritime institute revealed that they were disengaged with traditional assessments that comprised questions drawn from a question bank (Emad & Roth 2007). Over time, the students could predict the range of questions and prepare according to surface-learning approaches based on rote learning and memorisation. The ability to memorise is a lower-level cognition which is insufficient for performing in workplaces such as ships, where a higher level of cognition is required to assimilate, analyse and structure information for decision making and problem solving (Wiggins 1990). Providing memorised answers does not reflect the actual competence of the student to perform the same task on board a vessel (Lewarn 2002).

The STCW recommends that a seafarer's competence should not only be determined by the ability to integrate knowledge and skills in routine contexts but also by the ability to operate in unique and constantly changing conditions that require critical thinking and higher order cognitive skills (Walczak 1999). Seafarers who are trained to rely on memory and not on their ability to critically analyse the available information, may suffer from memory failure leading to human errors (Prasad 2011). Competence acquired through analysing and assimilating information enables professional students to learn how to select the correct course of action based on the gathered evidence and not purely on memory. This develops the seafarer students' professional competence to perform tasks based on the analysis of multitudes of information as experienced at the workplace.

1.2.2. Vagueness in the STCW Code

The SET institutes are expected to strictly comply with the requirements of the STCW Code by their respective national maritime regulators. However, even after the last revision of STCW in 2010, the vagueness in the STCW Code continues to leave too much room for interpretation by SET institutes, who use varying combinations of assessments (Bhardwaj 2009) for students to demonstrate the performance standards in the STCW Code. The STCW'95 did not fully eliminate the vagueness in assessment standards as it specified methods to demonstrate competence but did not provide specific methodologies, leaving it to the discretion of the assessor (Robson 2007). For example, from Table 1.2 it is evident that the STCW Code specifies methods (e.g. simulator training) to demonstrate competence but does not provide clear and detailed guidelines on how to use these methods to collect evidence of competence. For example, how sophisticated and advanced should the simulators be to reflect STCW standards since the STCW only provides recommended performance standards for non-mandatory types of simulators. In absence of descriptive and required standards for simulators, SET institutes may design assessment tasks on simulators that fail to assess the desired competencies.

In addition to non-descriptive simulators and methodologies, STCW also fails to provide unambiguous and clear terms for assessment of many units of competence. For example, Table 1.3 highlights words (promptly, minimize, etc.) from an extract in the STCW Code that does not provide clear terms of measurement. The words fail to provide a benchmark against which the SET institutes can assess how quickly, in terms of accuracy or completeness, the candidate should demonstrate the desired level of performance (Rutherford 1995). In the absence of a well-defined meaning, these words may be differently interpreted, leading to subjective assessment.

Table 1.3: Example of how the STCW Code fails to provide clear learning outcomes for seafarer competence assessment.

Competence	**Knowledge, understanding and proficiency**	**Methods for demonstrating competence**	**Criteria for evaluating competence**
Prevent, control and fight fires on board	Fire prevention and fire-fighting appliances Ability to organize fire drills Knowledge of classes and chemistry of fire…..	Assessment of evidence obtained from approved fire-fighting training and experience……	The type and scale of the problem is promptly identified….. Evacuation, emergency shutdown and isolation procedures are appropriate to the nature of the emergency…..

All the essential criteria for this task not identified and outlined

How soon is 'promptly'?

What does 'appropriate' mean and how can we measure it?

Performance standards should ideally communicate performance expectations from workplace duties, encompassing not only the technical skills but also the underpinning skills and knowledge. However, Table 1.2 provides an example of how the STCW code fails to recognise all necessary technical and underpinning skills and competencies required to perform workplace tasks. For example, planned distribution of cargo and recording information are not the only skills required for carrying dangerous goods. It should also identify essential skills, such as problem identification and determining whether there are any unexpected occurrences with its carriage. The SET institutes that complies strictly with the STCW Code assesses seafarer students in accordance with inadequate performance standards, producing graduate seafarers lacking adequate workplace skills.

The practice of assessing a limited range of skills can also curtail the development of a holistic portfolio of all the necessary skills required for supporting workplace performance at a particular level of responsibility (Cox 2009). It may be argued that the STCW Code does not specify the supporting skills in all units of competence, but can the seafaring industry afford to hide behind this oversight? Employers expect underlying competencies to be assessed along with the technical skills (Cross 2007). Competencies are skills that are distinct from technical and work-related skills that when used singularly or in various combinations integrally with technical skills, support the performance that defines competence (Teodorescu 2006). However, investigation of shipping accidents have often found that seafarers lack underlying competencies (Hetherington, Flin & Mearns 2006) and technical skills that allow them to develop the ability to transfer their competence to workplace scenarios.

1.2.3. Students lack ability to transfer skills to different contexts

A study by Sampson, Gekara & Bloor (2011) revealed that employers were unhappy with some of the current assessment methods that assessed a limited range of job- specific skills in settings that provide insufficient cues to the students on how the competence acquired in classrooms can be used in different contexts. The study comprised a series of 17 interviews with employers in the UK, the Philippines and Singapore that recruited seafarers on vessels involved in international voyages. The interviewees were fleet personnel managers from both owner-operator companies and ship management companies involved in the employment of seafarers largely from countries such as India, Myanmar, and the Philippines and from Eastern Europe. Similarly, official investigations and analyses of marine accidents have also revealed that seafarers assessed as competent in the use of particular skills in a given context failed to apply them in other contexts (Uchida 2004; Pecota & Buckley 2009; Prasad, Nakazawa & Baldauf 2010).

Although employers have training obligations for preparing their employees for specific types of vessels, the costs borne for aimless training should be avoided, as this can significantly affect employers' budgets. Seafarer employers need a reliable indication or contextual evidence of their employees' competence so that gaps in knowledge and skills can be identified and filled with additional training, if required. Assessments that do not provide contextual evidence may leave employers unsure regarding what the additional training should focus on. The International Safety Management (ISM) Code developed for the safe operation of ships clearly states that it is the responsibility of the seafarer employers to ensure their employees are competent to work on board ships (IMO 2002). The IMO authorises national regulators to investigate seafarers' competence through inspections and surveys, among many other regulatory requirements, to identify and deter substandard ships from operating (AMSA 2011). Ships can be detained, and registers cancelled if serious deficiencies are found in an operators' ability to perform workplace tasks (Department of Infrastructure and Transport 2012).

Successful performance in the workplace or during assessments in the SET institutes would require skills developed in a particular context to be transferred in different contexts and varied scenarios. However, the transfer of skills is affected by the context in which they were developed (Leberman, McDonald & Doyle 2006). There are skills defined by some authors (Clanchy & Ballard 1995) as transferable (or generic) skills that are not tied to particular contexts and are directly transferred to different scenarios. For example, a seafarer who has developed oral and written communication skills or the ability to plot a ship's position using data from the global positioning system (GPS) on hydrographic charts should be able to do so irrespective of the context or situation. On the other hand, there are non-transferable (domain specific) skills that are defined by some authors as skills tied to particular contexts, requiring students to learn how to use

them in different scenarios (Clanchy & Ballard 1995). For example, a seafarer that learns to competently manoeuvre a ship on a simulator may not be able to do so when given an actual ship. The assessment process can play a key role through the feedback provided on task performance. The feedback allows students and assessors to reflect on the application of their skills (Curry, Caplan & Knuppel 1994) in a particular context and identify additional training requirements for different contexts.

The outcome of seafarer assessments should be to inculcate such skills in students to allow them to make appropriate decisions in routine and unperceived situations and adapt to the diverse shipboard environment (Prasad 2011). To do so, students should be assessed via tasks in contextual scenarios that resemble workplace scenarios that produce sufficient and reliable evidence of the competence of seafarer students (Gonczi, Hager & Athanasou 1993). Although it may not be possible to recreate exact shipboard situations due to the complexity in structure and design, assessments designed to closely resemble workplace situations may provide cues for the transfer of competence for the student in the real world (McMullen & Braithwaite 2005). For example, the ability to demonstrate leadership skills may be assessed through a simulated emergency scenario that seafarer students may face on ships instead of a written assignment or pen-and-paper testing. This indicates that context similarity may be very relevant in seafarer assessment.

1.2.4. Traditional assessments lack validity and reliability to workplace contexts

Lack of contextual similarity in assessments that do not resemble workplace tasks raises the question of validity (relevance to the workplace) where students start to question the relevance of the assessments and the competence it purports to assess. For example, student interviews during the course review of navigation officers carried out in the Australian Maritime College (AMC) in 2011 revealed that seafarer students could not relate to multiple-choice questioning and oral examinations when it came to assessing competence for performing workplace tasks. They showed a preference for assessments that are contextually similar to the challenges found in the workplace in order to relate classroom learning to professional practice (AMC 2011).

Moreover, according to past researchers (Bailey 1998, 205; Law & Eckes 1995; Dikli 2003, p. 16; Abeywickrama 2012) traditional assessments have been conventionally described as not only inauthentic but also as a “one-shot” and single-occasion tests implemented at the end of a learning period, which makes them popularly classified under summative examinations. Hence, the scores obtained in the summative traditional assessments cannot inform on the progression of the learner as they only measure the students’ ability at a particular time (Law & Eckes 1995).

The use of summative traditional examinations at the end of the learning period represents the final judgement of the students' performance, and it is often too late to make any changes to the learning strategies. The student performance in the assessment tasks should allow valid generalisations about student competence (Wiggins 1992) with respect to the demonstrated learning outcome. However, such generalisation cannot be based on one performance, irrespective how complex or authentic the task was (Wiggins 1998). One of the ways generalisability across tasks may be achieved is to increase the number of performance assessments for each student providing them with more than one opportunity to demonstrate their mastery over the competence (Linn, Baker & Dunbar 1991).

Traditional assessments in seafarer education are usually summative and carried out at the end of the learning period, not allowing the students to engage in deep reflection during the assessment process. This is also the case with the oral examinations conducted to assess the seafarer's competence before issuing them with the certificate of competence (CoC). Seafarers who are unable to answer the questions to the satisfaction of the assessor are declared as 'fail' before being provided with another opportunity which often demoralises the students (Prasad, 2011). Implementing formative assessments would allow seafarer students to engage in metacognitive reflection to recognise the gaps that exist in their understanding. As gaps are recognized and become significant to students, they may locate, apply, and connect previous learning and new knowledge (Scott 2000) and skills causing the transfer of learning. Providing more than one opportunity with assessments that are contextually similar to the real-world assesses not only the seafarer students' ability to perform tasks to workplace standards (valid performance) but the ability to do so consistently (reliable performance).

To determine whether the intended learning outcomes (ILOs) have been achieved and to collate evidence, assessors need to decide whether the selected assessment methods adequately allow for the evaluation and demonstration of the ILOs (Moskal 2000). The quality of the information provided on outcome attainment is only as good as the assessments on which the reporting is based (Brindley, 1998). Thus, the ability to perform workplace tasks should be assessed through assessment methods that resemble professional scenarios. Hence, fidelity of context to conditions in which the professional skill would be applied becomes an important element of the adopted assessment methods. Such performance-based assessments applied in real-world contexts have often been described as authentic assessments (Herrington & Herrington 1998; Reeves & Okey 1996; Wiggins 1993; Meyer 1992).

1.3. Definition of authentic assessment

The idea of 'authenticity' in education was conceived and developed in response to increasing accountability to stakeholders. It was introduced in the 1980s in high schools in the United States of America. The term 'authentic' was first linked to student achievement by Archbald & Newmann (1988) requiring them to demonstrate outcomes beyond the school learning environment in an applied/work context. Wiggins (1989) related the term to student assessment, while promoting authentic assessment as a process that required student performance (Wiggins 1990) achieve standards expected in the professional area of practice. Unlike traditional tests that produced transcripts with ambiguous information of actual competence, evidence of student performance at workplace standards would improve accountability to stakeholders.

Authentic assessment is often used interchangeably with performance assessment as it involves some of the characteristics of the latter, but they are not synonymous (Marzano, Pickering & McTighe 1993). For example, all authentic assessments require a performance of some kind, but not all performance-based assessments are conducted in authentic or real-world contexts (Meyer 1992). Palm (2008) provided a detailed classification of meanings describing the similarities and wide range of differences between the meanings of each concept. Authentic and performance assessments are known as types of 'alternative assessments' to traditional assessments (Dikli 2003). Traditional assessments include pen-and-paper testing, multiple-choice questioning, and oral examinations.

Cumming & Maxwell (1999) showed that characteristics of authentic assessment can also be found in other assessments, such as problem-based and competency-based assessments but provided a clear distinction between them. For example, they explained that authentic assessment is based on theories of learning where task performance occurs in a genuine workplace or contextually similar situations. On the other hand, competency-based assessments are based on the theory of vocational education where assessment tasks should represent workplace tasks but can be performed in individual components and are not necessarily integrated into one holistic task. Authentic assessments have also been called dynamic assessments (Chance 1997; Butler 1999) due to the dynamic nature of evolving to address student learning needs.

In the field of education, authentic assessment is typically presented as a model that requires students to provide responses to a situation described and delivered in real-world (or contextually similar) contexts (Villarroel et al. 2018). Authentic assessment tasks are meaningful to students due to its strong figurative context and fidelity to the situations that the students may find themselves in within the professional world (Wiggins 1989; Gulikers 2006). Meaningful tasks set in real-world contexts enhance

student engagement with assessment if students relate the tasks to professional practices (Richards-Perry 2011; Quartuch 2011).

1.4. Research gaps

Although authentic assessments are applied in real-world settings to ensure the assessments have a high fidelity to real-world contexts and require competence as expected at the workplace, assessments should be judged for its technical adequacy of measures by the established criteria of validity and reliability (Linn et al. 1991). Validity refers to the extent to which the evidence produced through assessments supports the inferences made about the students' competencies and whether such inferences are being interpreted in appropriate contexts; and reliability refers to the consistency of assessment scores obtained every time the same competencies are assessed irrespective of the scorer, period between the assessments, and contextual and individual learning variables under which the assessments occur (Moskal & Leydens 2000).

Reliability and validity problems are found to be very typical of authentic assessment. It is often assumed that reliability is achieved concurrently with validity, and thus may be ignored or accepted with low levels in traditional assessments, as evident in the study by Olfos & Zulantay (2007). Thus, reliability is often accepted as a necessary condition of validity. However, in cases of authentic assessment, reliability cannot be ignored or accepted with low levels as a trade-off between validity and reliability (Jonsson 2008). Reliability mainly indicates consistency of performance which is essential for workplace-based tasks.

One of the ways to assure validity and reliability of authentic assessments is through the use of assessment rubrics. Rubrics combine the essential criteria and the levels of performance by which the performance would be judged (Mueller 2005). The criteria and the levels forms a scoring guide for the assessment making it easier for educators to define what is being measured through the assessments and how the score should be interpreted (Emery 2001).

There are two main methods of measuring the performance of a student in an assessment: objective and subjective measures (Nazir, Jungefeldt & Sharma 2019). Authentic assessments are largely subjective assessments which may have more than one correct answer for scoring purposes as opposed to objective assessments which have a single and defined solution to designed problems (Nazir et al. 2019). Scoring without specific standards or guidelines to inform assessment and evidence gathering may lead to subjective judgements. Rubrics are an option that can be used to improve the fairness in scoring by specifying the same criteria and standards to be applied to all students'

work for scoring by either individual or multiple assessors (Dennison, Rosselli & Dempsey 2015).

Reliability and validity are crucial to the acceptance of authentic assessment as an accurate measure of knowledge, skills, and behaviours (Stevens 2013). There are numerous extraneous variables that affect the validity and reliability of the rubrics when used as an assessment instrument (Taylor 2011). If these variables are not addressed, then the validity and reliability of the assessment and the resulting outcomes become questionable (Olfos & Zulantay 2007). For example, a lack of construct validity may indicate that the underlying psychological variables such as problem-solving, social interaction, and communication which are required universally in most professions, were not adequately assessed in these cases. In the case of seafarer education, improving validity and reliability may enhance students' perception of authenticity in assessments, and as a result, lead to higher engagement and ability to transfer skills to workplace scenarios.

Hence, to find the research gaps, this research studied the past approaches to validity and reliability of authentic assessments; and whether valid and reliable authentic assessments have been implemented in seafarer education. This section identifies the two research gaps that guided the research presented in this thesis.

1.4.1. Lack of a holistic approach to validity and reliability of authentic assessment

Based on an extensive literature review conducted for this research study (as explained in Chapter 2), the first gap related to the extent of investigation conducted in the area of validity and reliability of rubrics as authentic assessment tools. Building on and extending the first literature review, a second literature review (explained in Chapter 2) also included literature that discussed implementation of authentic assessment without the use of assessment rubrics. The review of the selected articles reflected the absence of both validity and reliability testing or addressing only few aspects of it.

Hence, a novel contribution of this research study was made through the development of a conceptual framework that specifically addressed on the different aspects of validity and reliability associated with authentic assessment

1.4.2. Absence of research on authentic assessment in seafarer education

Past research (Brawley 2009; Schneider et al. 2002; Thomas 2000; Leon & Elias 1998; Gallagher, Stepien & Rosenthal 1992) has empirically proved higher student academic achievement for authentically assessed students when compared with their traditionally

assessed counterparts. However, the literature reviews conducted as part of this research study (Chapter 2), revealed that similar evidence was essentially missing in the area of seafarer education.

The extensive literature reviews also revealed empirical evidence to prove that the implementation of authentic assessment has enhanced student engagement (Richards Perry 2011; Quartuch 2011; Findlay 2013) and their ability to transfer learning (Herrington & Herrington 1998; Sator 2000; Saunders, Saunders & Batson 2001) in different contextual scenarios. Past research (Bell & Bell 2003; Cassidy 2009; Wellington et al. 2002) has also shown that authentic assessment has been implemented to investigate its effect on the achievement of educational or professional standards, constructive alignment of instruction processes with assessment, and achievement of professional competence (including demonstration of essential behaviours). Similar research is needed but has been largely ignored in the area of seafarer education.

Hence, a novel aspect of this research study was the collection of empirical evidence regarding the impact of authentic pedagogical practices on seafarer students' academic achievement.

1.5. Research questions

As discussed previously, past research has shown that traditional assessments that require seafarer students to focus on rote learning and construction of responses devoid of context resulted in the students' disengaging from learning. Memorising information is a lower-order cognitive ability, the failure of which has led to errors in assessment tasks resulting in low academic achievement for students. Authentic assessment presents a model that requires students to construct responses through the critical analysis of information presented in real-world contexts. Consistent with the previous discussions, the main interest of this study was comparing seafarer students' academic achievement in authentic assessment as compared with traditional assessment. Since some past researchers (Bailey 1998; Law & Eckes 1995; Dikli 2003; Abeywickrama 2012) defined and implemented traditional assessments as a summative assessment, this project implemented the assessment in a summative format as opposed to authentic assessments implemented formatively.

This section formulates a set of research questions (RQ) addressing the previously identified problems and gaps.

RQ1: Is there a significant improvement in seafarer students' academic achievement in authentic assessment when the scores are compared with traditional assessment scores?

RQ2: Is there a significant improvement in seafarer students' academic achievement in formative authentic assessment when the scores are compared with summative traditional assessment scores?

Taking the investigation of RQ1 and RQ2 forward, this research also aimed to investigate the seafarer students' perception of authenticity with factors of assessment (task, criteria, etc.) in the designed assessment task. The students' perception of authenticity, if correlated with their academic achievement (scores) was intended to explain factors that students perceived to be significant towards their academic achievement. Hence, the following RQ was formulated:

RQ3: What is the correlation between seafarer students' perception of authenticity in assessment and their academic achievement in the associated assessment tasks?

1.5.1. Linking research questions to the published work

This thesis forms a capstone to previous, extended investigations conducted within this PhD research project from March 2013 to May 2019. In particular, the doctoral work contains seven fully refereed papers (publications) prepared during the candidature period. The research presented in these papers as well as the respective findings are synthesised in Chapter 5, and the actual publications are reproduced in Chapter 7. Five of the publications have been published and two are under review.

This capstone thesis is designed to demonstrate that the previously mentioned publications constitute essential parts of a coherent and integral body of work. Table 1.4 outlines the research focus of the seven publications included in the thesis and shows how they were related to a single research project. As indicated in Table 1.4, the seven publications are linked thematically to the research problems identified in Section 1.2 (1.2.1—1.2.4), the research gaps presented in Section 1.3 (1.3.1—1.3.2), and to the research questions formulated in Section 1.4.

Table 1.4: Linking the publications to the research problems, gaps and questions.

	Research focus	Problems, gaps, and questions investigated	Related papers
1.	Why is authentic assessment required for seafarer education?	1.2.1 1.2.3 1.2.4	Paper I
2.	Why does the STCW Code fail to provide outcomes for the authentic assessment of seafarer students?	1.2.2	Paper II
3.	How can authentic assessments enhance student engagement and the ability to transfer competence to different contextual scenarios?	1.2.1 1.2.3	Paper III
4.	How was the validity and reliability of authentic assessment addressed in the past research?	1.3.1	Paper IV
5.	How can the validity and reliability of authentic assessment be addressed holistically?	1.3.1	Paper V
6.	Is there a significant improvement in seafarer students' academic achievement in authentic assessment when compared with traditional assessment?	1.3.2 1.4 (RQ1 and RQ2)	Paper VI
7.	What is the correlation between seafarer students' perception of authenticity in assessment and their academic achievement?	1.3.2 1.4 (RQ3)	Paper VII

1.6. Outline of the thesis

After having established the research background, problem definition, research gaps, and research questions in Chapter 1, the remainder of this thesis is organised as follows.

Chapter 2 explains the literature review conducted for the purpose of finding research gaps in the area of authentic assessment.

Chapter 3 discusses the theoretical framework (Section 3.1) of this study and, in particular, the underlying theories to achieve the authentic assessment outcomes (Section 3.2) of higher student engagement (Section 3.2.1); ability to transfer skills to different contextual scenarios (Section 3.2.2); contextual and multiple examples of evidence of competence (Section 3.2.3); and valid and reliable student performance (Section 3.2.4). The chapter then develops a novel and practical conceptual framework to improve the different aspects of validity and reliability of authentic assessment when implemented in the context of seafarer education (Section 3.3.).

Chapter 4 provides details on the methods and materials used in this research, specifically on the research methodology chosen for the study (Sections 4.1 and 4.2). The

approach selected to collect (Sections 4.1.1 and 4.2.3) and analyse the quantitative data (Sections 4.1.2 and 4.2.5) is also discussed.

Chapter 5 focuses on the results published in the framework of this research. The findings reported in the respective publications are summarised in Sections 5.1 and 5.2.

Chapter 6 discusses the research findings and their implications (Section - 6.1), research contributions (Section 6.2), limitations and constraints (Section 6.3), future research (Section 6.4), and the conclusion (Section 6.5) of this study.

Chapter 7 reproduces each of the papers (publications) included in this thesis. The outline of the thesis is summarised in Table 1.5.

Table 1.5: Outline of the thesis.

Chapter number and title	Sub-topic	
1. Chapter 1: Introduction	1.1	Background
	1.2	Problem definition
	1.3	Definition of authentic assessment
	1.4	Research gaps
	1.5	Research questions
	1.6	Outline of the thesis
2. Chapter 2: Literature review	2.1	Validity and reliability of authentic assessment
	2.2	Search and selection of articles
	2.3	Research gaps
3. Chapter 3: Theoretical and conceptual framework	3.1	Theoretical framework
	3.2	Underlying theories for achieving authentic assessment outcomes
	3.3	Conceptual framework
4. Chapter 4: Research methodology	4.1	Research methodology – Part 1
	4.2	Research methodology – Part 2
5. Chapter 5: Results	5.1	Results for RQ1 and RQ2
	5.2	Results for RQ3
6. Chapter 6: Discussion and Conclusion	6.1	Research findings and their implications
	6.2	Research contributions
	6.3	Limitations and constraints
	6.4	Future research
	6.5	Conclusion
7. Chapter 7: Appended papers	7.1	Paper I
	7.2	Paper II
	7.3	Paper III
	7.4	Paper IV
	7.5	Paper V
	7.6	Paper VI
	7.7	Paper VII

Chapter 2 explains the literature review conducted for the purposes of finding the research gaps in the area of authentic assessment. The first literature review was conducted to investigate past approaches to validity and reliability of authentic assessment, when implemented with assessment rubrics. The second literature review extended the first review to include additional research when authentic assessment was implemented without assessment rubrics. The second review also investigated the existence of a conceptual framework that addresses the validity and reliability of authentic assessment. In terms of publications included in this thesis, the first review was the basis of Paper IV and the second review was the basis of Paper V.

2.1. Validity and reliability of authentic assessment

2.1.1. Different types of validity and reliability

Authentic assessment is defined as assessments resembling real-world contexts. However, the fidelity of context alone cannot ensure that essential aspects and constructs of professional competencies are being accurately assessed. Advances in technology, such as simulators, web-learning, multimedia, etc. have allowed many researchers (Neely & Tucker 2012; Neo, Neo & Tan 2012; Scholtz 2007) to use such technology to create authentic experiences that can replicate real-world tasks for students. However, Messick (1996) was not convinced that authentic assessments can ever fully represent real-world tasks in educational settings and suggested that the assessments are prone to threats of validity, which emphasises the appropriateness of assessment tasks as effective measures of ILOs (Rhodes & Finley 2013).

In the area of education, validity is not seen as a property of the assessment but how the results have been interpreted (Jonsson & Svingby 2007). Validity refers to the degree to which evidence produced from assessments supports the interpretation of a student's competencies. Table 2.1 describes the three types of evidence that are commonly examined to support the validity of an assessment instrument: content, criterion, and construct (Moskal & Leydens 2000).

Table 2.1: Types of validity testing in student assessments.

Validity		
Content Validity: extent to which the assessment instrument provides a representative sample of the content domain in the area of interest (Lynch 2003).	**Criterion Validity:** extent to which a student's performance on a test accurately predicts the student's performance on an external criterion (Lynch 2003).	**Construct Validity:** Extent to which the assessment measures the theoretical construct or processes that are internal to an individual (Moskal & Leydens 2000).

It is extremely difficult to construct an assessment that is truly valid in measuring what it is supposed to measure (Finch 2002). For example, an assessment designed to assess a student's ability to fight fires may not be able to effectively measure personal or professional behaviours (such as creativity and critical thinking) associated with the task performance. According to Messick (1996), it is hard for assessments to achieve complete validity but the threats to validity can be minimised by ensuring that assessments do not contain anything that is irrelevant to the measurement of the desired outcomes. For example, assessments designed to assess a student's ability to fight fires should not include pen and paper testing in classrooms which are irrelevant to the measurement of either the task performance or behaviours associated with it.

Capturing a more authentic performance does not ensure validity (Stevens 2013). For example, Hoepfl (2000) pointed out that creating standards for authentic assessments is a challenging task that may suffer from 'construct underrepresentation' if the standards fail to assess essential dimensions of knowledge and skills or 'construct-irrelevant variance' if the standards require tasks that are not relevant to measuring the desired competencies (Messick 1995). Assessments are valid if they effectively measure the ILO they were designed to assess. Whether assessments effectively measure the ILOs cannot be based on the subjective judgement of whether the questions appear to do so, which is known as face validity (Drost 2011). Drost (2011) explained that, although face validity is important for credibility to stakeholders, it is the weakest and least scientific form of establishing validity for assessments.

Ideally, an assessment should produce similar results independent of the scorer and the context of assessment. The more consistent the scores are over different scorers and contexts, the more reliable the assessment is thought to be. Table 2.2 provides the different types of reliability testing conducted in the area of education.

Table 2.2: Types of reliability testing in student assessments.

Reliability			
Inter/Intra Rater Reliability: variations in judgements across raters/markers (inter-rater) or in the consistency of one single rater/marker (intra-rater) (Jonsson & Svingby 2007).	**Test-retest Reliability:** consistency of results when the same test is administered after a specific period (Drost 2011).	**Split-half Reliability:** two tests and two measures assessing the same construct (Drost 2011).	**Internal Consistency Reliability:** how well the different components of the assessment measure a particular construct (Drost 2011).

Methodologically sound assessment instruments should have acceptable levels of both validity and reliability (Rhodes & Finley 2013).

2.1.2. Achieving validity and reliability through assessment rubrics

Rubrics (example shown in Appendix 8) are assessment tools that comprise individual and essential dimensions of performance known as criteria along with standards for levels of performance against those criteria (Jonsson & Svingby 2007). Although the terms 'criteria' and 'standard' are sometimes used interchangeably, they have distinct meanings (Sadler 2005). The definitions provided by Sadler (2005) and Spady (1994) provide a robust basis for distinguishing the terms. Standards are defined as levels of definite attainment and sets of qualities established by authority, custom, or consensus by which student performance is judged, whereas criteria are essential attributes or rules used for judging the completeness and quality of the standards.

Moreover, OBE, such as seafarer education, requires the ILOs be established and communicated to students beforehand to make the learning process transparent (Biggs & Tang 2007). As assessment rubrics communicate standards and feedback for achievement, they are an essential tool for OBE (Reddy 2007). Rubrics provide clear statements on learning and performance expectations for both educators and students. Such statements can then be used to assess whether the intended outcomes were achieved by students, educators, and assessors. Hence, rubrics are highly regarded as tools that increase validity and reliability in assessments (Rezaei & Lovorn 2010; Jonsson & Svingby 2007; Silvestri & Oescher 2006).

The validity of the results and strength of the rubric as an assessment instrument are evidenced by positive results on a variety of reliability tests (Diller & Phelps, 2008). For example, according to Jonsson & Svingby (2007), one widely cited effect of rubrics in the areas of authentic and performance-based assessments is the consistency of judgement and scoring across students, tasks, and different raters (scorers); and Vendlinski et al. (2002) used rubrics to authentically assess 134 first-year high school chemistry students

to achieve valid inferences of a student's content understanding, while not allowing the score to be affected by gender, ethnic, or socioeconomic bias. Without rubrics, students have no guidelines towards achievement or to understand the teacher's feedback comments (Montgomery 2002) on achieved outcomes. For example, using a focus group discussion involving 14 undergraduate students, Andrade & Du (2005) found the use of rubrics to be very effective in providing performance expectations and feedback about the achievement of standards in teacher education. Rubrics may report on outcomes attainment, but the validation of attainment is achieved through the assessment process (Davis et al. 2007).

Using rubrics to communicate standards achieved by students in professional education also requires assessment methods such as authentic assessment that can capture such standards. Traditional assessments assess the ability to recall facts and some of the applied skills (Archbald 1991) but fail to assess essential behaviour-based attributes (Wiggins 1992) that an individual must develop along with technical skills and knowledge that together define professional competence (Sampson & Fytros 2008). Assessment of professional competence can be captured through authentic assessment tasks that are based on meaningful contexts and applied in real world or contextually resembling real-world settings. However, professional competence is developed and assessed under specific contexts in educational settings. Transfer of performance or competence to perform individual components of a task to a holistic performance of the task where integration of competence is required cannot be assumed (Cumming & Maxwell 1999).

According to Cumming & Maxwell (1999), learning and assessment needs to be contextualised to make it relevant and meaningful for students. Meaningful context can provide not only motivational benefits to student learning but also a clear understanding of learning that can or cannot be transferred to different contextual scenarios. If real-life contexts and complexities (task-centred approach) cannot be created in assessments, they should then focus on the selected constructs (construct-centred approach) of knowledge and skills (Messick 1996). For example, assessments designed in SET may not be able to assess a student's competence to manage large crowds as is required on passenger ships, but they may be designed to assess a student's competence to do so through their ability to analyse risks associated with such management or to develop crowd management plans. Although such assessments may take place in controlled situations, the authenticity is reflected through the ways in which the same skills would be applied in real-life contexts (Messick 1994). The standard of learning achieved in the real-world contexts may be communicated via rubrics making it an important authentic assessment instrument for assessing outcomes that represent workplace tasks.

One of the key characteristics requires authentic assessment to provide performance criteria to students beforehand, which can be done through the use of rubrics. Provision

of clear expectations of standards of performance via rubrics allows students to learn and educators to adopt appropriate instructional strategies to guide students towards the achievement of the desired outcomes (Archbald 1991). The use of summative traditional examinations at the end of the learning period represents the final judgement of the students' performance and it is often too late to make any changes to the learning strategies. Authentic assessment methods that are based on the ongoing use of formative assessments may be more suitable to provide diagnostic feedback and make adjustments to improve the learning process (Burke 2011). Hence, the alignment of the learning, teaching, and instruction process towards the achievement of outcomes creates constructive alignment (Biggs & Tang 2007).

Constructive alignment comes from the constructivist theory (Biggs & Tang 2011), where the student is not a mere receiver of knowledge but is also actively involved in the construction of it while progressing in learning. Newmann, Marks & Gamoran (1996) and Cooperstein & Kocevar-Weidinger (2004) connected authentic assessment to constructivist learning. Although principles of constructivism can allow everyone to construct meaningful learning, Newmann et al. (1996) recommended that high intellectual standards provided through rubrics in authentic assessment can promote highly intellectual construction of knowledge and meaning, leading to superior learning and performance that would require students to use higher-order cognitive skills.

Performance-based assessments like authentic assessment face the problem of obtaining reliability (Lynch 2003). Issues such as lack of reliability, inconsistency in assessment design and grading, and potential for grading bias remain important challenges with authentic assessment (Rhodes & Finley 2013). Addressing and improving on different aspects of validity and reliability provides evidence of a student's ability to perform assessment tasks using real-world competencies to workplace standards and to do so consistently, ensuring a holistic approach to competence assessment.

Authentic assessments represent real-world tasks as valid indicators of workplace competence which should be consistent irrespective of the context or scorer. Such consistency can only be proved through reliability. Hence, authentic assessments should achieve both validity and reliability. Because it can be difficult to establish whether an assessment instrument truly captures the outcome for which it is intended or whether the outcome can be consistently measured, it is preferable for instruments to demonstrate more than one type of validity (Rhodes & Finley 2013) and reliability. There are numerous aspects of validity and reliability investigated and reported in the literature on assessment. They may be discussed selectively, but none should be ignored (Jonsson & Svingby 2007). Although rubrics do not make assessment valid, addressing different aspects empirically could make assessments more valid and reliable for the intended purpose, eliciting the required performance (Jonsson 2008). There is sparse research

focussing on the quality of rubrics as a valid and reliable assessment tool (Stellmack et al. 2009). Hence, a literature review in the area of authentic assessment was carried out to reveal whether a holistic approach to improving its validity and reliability through rubrics (and without it) has been used by past researchers.

2.2. Search and selection of articles for the literature review

The first literature review (discussed in Paper IV) analysed 124 articles, spanning from 1989 (when authentic assessment was first introduced) to 2015 (when Paper IV was written and submitted for publishing). However, the review was updated in 2019 before this thesis was submitted for examination. In the period between 2015 and 2019, an additional 112 articles were reviewed (taking the total number of articles to 236).

The second literature review built on and extended the first review. In addition to the 124 articles (from the first literature review), an additional 28 articles (total = 152 articles) were analysed to investigate the existence of a conceptual framework that has a holistic approach to the validity and reliability of authentic assessment. In doing so, it also investigated past approaches to validity and reliability of authentic assessment when implemented with and without assessment rubrics. The articles spanned from 1989 (when authentic assessment was first introduced) to 2016 (when Paper V was written and submitted for publishing). However, the review was updated in 2019 before this thesis was submitted for examination. In the period between 2016 and 2019, an additional 84 articles were reviewed (taking the total number of articles to 236).

The articles reviewed included books, chapters in books, conference papers and proceedings, government documents, journals, reports, theses, and other articles classified as generic. The articles were chosen after a web-based search on popular websites such as Google, Google Chrome, and Google Scholar as well the library database of the University of Tasmania to assure adequate breadth and depth of coverage.

The University of Tasmania uses popular search systems and databases, such as ProQuest and Web of Science which enabled widening the search for articles. The database of ProQuest was used to identify scholarly journals, conference papers and proceedings, dissertations and theses, report and other evidence-based resources that included a discussion on authentic assessment. The search engine of the Web of Science database was used to enable more complex citation checking and evaluation of scholarly articles as compared to other search engines such as Google Scholar. In addition to searching for articles, the Web of Science database also categorised the identified articles into disciplines, sub-disciplines and other scholarly domains which was found useful for the purpose of organising the literature review. Articles were also found using the snowballing technique based on a search through citations in articles discovered through

an online search. An effort was made to obtain as many articles as possible through the previously defined methods.

Peer-reviewed articles were sought using the abstract search of the ProQuest and Web of Science with the keywords and Boolean operators:

1) “authentic assessment” OR “authenticity in assessment” OR “authentic” OR “authenticity” OR “performance assessment”

AND

2) “seafarer education and training” OR “engagement” OR “transfer” OR “validity” OR “reliability” OR “evidence of competence” OR “rubrics” OR “student performance”.

The first set of keywords reflect those used in the main literature review conducted in the field of authentic assessment by past researchers (Ashford-Rowe 2009; Palm, 2008; Taylor, 2011; Varley 2008). The second set of keywords was used to identify published research that investigated the relationship between authentic assessment and the outcomes of engagement, transfer of learning, evidence of competence, and valid and reliable student performance outcomes. Hence, all reviewed articles contained both the words ‘authentic’ and ‘assessment’ or ‘authenticity’ and ‘assessment’ in their titles. However, there were a few exceptions (e.g. Wiggins 1998 and BoarerPitchford 2010) when the focus of the article was centred around the topic of authentic assessment.

2.2.1. Criteria for the inclusion and exclusion of articles for the literature review

The purpose of the literature review was to highlight the different types of validity and reliability demonstrated in past research, when authentic assessment was implemented for the testing of students’ knowledge and/or skills with the use of assessment rubrics (first literature review) and without the use of it (second literature review). Hence, scholarly articles obtained through the database search were reviewed to search for a discussion on the same purpose, and if found, was selected for the review. The selected articles that addressed at least one aspect of validity and reliability in its research were also included to review for the investigation of an existing conceptual framework.

Scholarly articles that included a theoretical discussion (literature reviews, theoretical models/frameworks) and empirical research works that were not based on the implementation of assessment were excluded from the literature review. Table 2.3 provides a snapshot of the criteria used for the inclusion and exclusion of articles for the literature review.

Table 2.3: Criteria used to select articles for the literature review.

Total number of articles reviewed	**236**
Articles **excluded** based on the non-implementation of authentic assessment (includes theory discussion, literature reviews, theoretical models/frameworks, empirical papers that excluded assessment implementation).	**177**
Articles **included** based on the implementation of authentic assessment but without the use of assessment rubrics.	**23**
Articles **included** based on the implementation of authentic assessment with the use of assessment rubrics	**36**
Articles **selected** from the above included articles (based on the addressing of at least one aspect of validity and reliability) to investigate the use of a conceptual framework	**14**

2.3. Research gaps from the literature review

2.3.1. Research gaps from the first literature review

Table 2.3 shows that a total of 236 articles were analysed in literature review 1. 177 articles were conceptual in nature and hence, excluded from the analysis. 36 articles were selected for the first review since they discussed the implementation of authentic assessment with the use of assessment rubrics (summarised in Appendix 1).

The intention of the literature review was to find the extent of investigation that has been carried out in the area of testing validity and reliability of rubrics as authentic assessment tools. The gap found in this respect reflected the absence of both validity and reliability testing in some studies such as Todorov and Brousseau (1998), Emery (2001), Vendlinski et al. (2002), and Brawley (2009). The review revealed that past research in the area of authentic assessment has addressed typically only one or two aspects of validity and reliability while others have not been investigated. Barring one study by Jonsson (2008), none of the studies in the classification demonstrated construct validity. A lack of construct validity may indicate that that underlying psychological variables such as problem-solving, social interaction, and communication which are required universally in most professions were not adequately assessed in these cases.

Some studies revealed other types of validity, such as face and convergent validity which were not categorised under the three common types of evidence required to support the validity of an assessment instrument. While face validity is the weakest and least scientific form of establishing validity; convergent validity was explained by Cassidy (2009, p. 106) as a subcategory of construct validity that seeks “agreement between a theoretical concept and a specific measuring instrument”. The review revealed that some researchers like Cassidy (2009) use a pre-tested instrument expecting the same validity and reliability as obtained in previous studies. However, if using a pre-existing

instrument, it is essential for researchers to establish the instrument's validity and reliability in the context of their own research (Burton & Mazerolle, 2011).

According to Lovorn & Rezaei (2011), simply using rubrics does not improve the reliability of the assessment. Reliability can only be improved if rubric users are well-trained on its development and use. Raters/scorers need to be involved in the development of rubrics because it takes time for them to understand its purpose and implementation (Diller and Phelps, 2008). For example, the study by Lovorn & Rezaei (2011) involved the training of 55 teachers in rubric use to find a resulting increase of reliability in writing assignments. However, many of the studies such as those by Moon et al. (2005), Olfos & Zulantay (2007), Diller & Phelps (2008), do not mention any training for rubric users before they were administered. In the study by Taylor (2011), teacher development workshops were carried out to minimise threats to internal validity only. However, according to Taylor (2011), training conducted for rubrics development or use should be consistent for all involved. Differing approaches in terms of context, standards, or application can influence the results of research data and create problems with validity.

The review also revealed that although past investigations in authentic assessment and its influence on learning were conducted in the contextual settings of school, vocational, and university education, there was an absence of research on authentic assessment in the field of seafarer education.

The first review revealed the absence of a holistic approach to address and improve on the different aspects of validity and reliability of authentic assessment. This warranted an extensive search for an existing conceptual framework in the field of authentic assessment that could serve the purpose. However, a key limitation of the first review was the exclusion of literature that discussed the implementation of authentic assessment without the use of rubrics. To do so, the second literature review was required. The second review included the investigation of past approaches to validity and reliability of authentic assessment when it was implemented without assessment rubrics. With these objectives, this research conducted a second literature review to reveal further research gaps in the area of authentic assessment.

2.3.2. Research gaps from the second literature review

Table 2.3 shows that a total of 236 articles were analysed in the second literature review. A total of 23 articles discussed the practical implementation of authentic assessment without the use of assessment rubrics. The analysis of these 23 articles (as summarised in Appendix 2) revealed that barring one, none of the papers addressed any of the aspects of the validity and reliability of authentic assessment.

The first and the second literature review revealed that a total of 14 articles addressed one or two aspects of validity and reliability when authentic assessment was implemented with and without assessment rubrics. This was summarised in Appendix 3. The second review also found the absence of a global conceptual framework (as highlighted in Appendix 3) that identifies and practically addresses the different aspects of validity and reliability of authentic assessment, justifying the need to develop one especially in the context of seafarer education. Similar to the first review, the second review also revealed an absence of research on authentic assessment in seafarer education.

The next chapter (Chapter 3) explains the theoretical framework formulated by redefining the concept of authentic assessment. It also explains the conceptual framework constructed in this thesis to practically address the different aspects of validity and reliability of authentic assessment when implemented in seafarer education.

3. THEORETICAL AND CONCEPTUAL FRAMEWORK

This chapter explains the theoretical framework adopted in this research and the underlying theories that underpin the framework. The definition of authentic assessment (derived from the literature review) provided the theoretical framework for this research. The underlying theories of constructivism, self-efficacy, and metacognitive reflection underpin the framework. This chapter also develops a novel conceptual framework that practically addresses the different aspects of validity and reliability of authentic assessment when implemented in seafarer education. In terms of publications included in this thesis, the theoretical framework is discussed in Paper IV, the underlying theories that lead to authentic assessment outcomes are discussed in Paper III, and the conceptual framework is developed in Paper V.

3.1. Authentic assessment redefined (Theoretical Framework)

Past researchers have employed different methods of authentic assessment such as portfolios, journals, case studies, observations, workplace assessments, report writing, self- and peer-assessment, group work and presentations. This research used the extensive literature review (explained in Section 2.2) to define authentic assessment by collating the characteristics provided by the most commonly cited authors (e.g. Wiggins 1989; Archbald 1991; Darling-Hammond & Snyder 2000) in the area (see Table 3.1). The exact number of citations for the individual papers has been obtained from the Google Scholar website. The authors listed in Table 3.1 conducted their research in different educational contexts such as high or elementary schools, teacher education, and nursing education.

Table 3.1: Characteristics of authentic assessment defined by the most commonly cited authors.

Most commonly cited author/Year of publication	No. of citations	Context of Study	Real world context	Integration of competence	Known performance criteria	Valid and reliable performance	Forward looking questions	Ill-structured problems	Evidence collected by student	Tasks resulting in outcomes	Contextual evidence of competence	Multiple indicators of competence	Promote student engagement	Allow transfer to different contexts
Wiggins 1989	939	High Schools	✓	✓	✓	✓	✓	✓	✓			✓		
Wiggins 1990	383	High Schools	✓	✓	✓	✓	✓	✓	✓			✓		
Wiggins 1993	364	High Schools	✓	✓	✓	✓	✓	✓	✓			✓		
Wiggins 1998	956	High Schools	✓	✓	✓	✓	✓	✓	✓			✓		
Archbald 1991	27	Schools			✓					✓			✓	✓
Darling-Hammond & Snyder 2000	397	Teacher Education	✓	✓	✓							✓		
Gulikers, Bastiaens, & Kirschner 2004a	212	Nursing Education	✓	✓	✓					✓	✓			
Gulikers, Bastiaens, & Kirschner 2004b	17	Nursing Education	✓	✓	✓					✓	✓			
Gulikers, Bastiaens, & Kirschner 2006	33	Nursing Education	✓	✓	✓					✓	✓			
Gulikers, Bastiaens, Kirschner, & Kester 2008	23	Nursing Education	✓	✓	✓					✓	✓			
Gulikers 2006	10	Nursing Education	✓	✓	✓					✓	✓			

Based on the characteristics provided in Table 3.1, authentic assessment herein will encompass the following:

- **Tasks** *resulting in outcomes* in a *real-world context* that require the *integration of competence* to solve *forward-looking questions* and *ill-structured problems*,
- **Processes** that require *performance criteria to be provided beforehand* and *evidence of competence to be collected by the student*, and
- **Outcomes** that result in *valid and reliable student performance, contextual and multiple examples of evidence of competence, higher student engagement, and transfer of skills to different contexts*.

The definition is presented in a tabular format in Table 3.2.

Table 3.2: Authentic assessment redefined based on characteristics provided by the most commonly cited authors.

Authentic Assessment		
Tasks	**Processes**	**Outcomes**
• Set in a real-world context • Requiring an integration of competence • Comprising of forward-looking questions • Ill-structured problems	• Requiring performance criteria to be provided beforehand • Evidence of competence to be collected by the student	Resulting in: • Higher student engagement • Ability to transfer skills to different contexts • Contextual and multiple evidence of competence • Valid and reliable student performance

Thus, the definition of authentic assessment, where its implementation results in the previously stated outcomes (Table 3.2), provided the theoretical framework for this research. However, to ensure that the 'authentic' tasks reflect workplace situations requiring students to apply knowledge, skills, and behaviours to professional standards and to test the consistency of such performances, it has already been established that authentic assessments and the resulting performances should be judged by the essential criteria of validity and reliability. Hence, using theories of learning (constructivism, self-efficacy, and metacognitive reflection) and empirical evidence from past research, the next section justifies the use of the theoretical framework for the purposes of this research. It focuses on how addressing and improving the different types of validity and reliability are essential for authentic assessment to achieve its outcomes in seafarer education.

3.2. Underlying theories for achieving authentic assessment outcomes

3.2.1. Outcome: Higher student engagement

Authentic assessment requires tasks to resemble real-world scenarios or similar contexts. Real-world scenarios provide meaningful contexts for knowledge and skill application for students, thus creating a high level of student engagement and commitment (Richards Perry 2011; Pallis & Ng 2011). However, how do we ensure that the authentic tasks designed by the educators are perceived by the seafarer students as valid and relevant to workplace tasks?

Content validity evaluates the extent to which the assessment instrument provides a representative sample of the content domain in the area of interest (Lynch 2003). For example, if the authentic assessment were designed to assess a seafarer student's competence to fight fires on board a ship, content validity of the assessment would ensure that it adequately covers the content of fire-fighting practices and conditions on ships. It would also ensure that the assessment does not contain anything that is irrelevant to the measurement of the ability to fight fires. Hence, content validity is popularly achieved through validation by subject experts (Oh et al. 2005; Lang II 2012). However, it is a rational analysis based upon individual, subjective judgement (Moon et al. 2005), which may result in bias. The bias may be reduced if multiple subject experts are employed for validation. (Moon et al. 2005). Table 3.3 highlights the relevance of content validity testing to authentic assessment.

Table 3.3: Relevance of content validity testing to authentic assessment.

Objective	Validity tested	Achieved through
How do we ensure authentic tasks designed by the educators are perceived by the seafarer students as valid and relevant to workplace tasks?	Content validity	Validation by multiple subject experts

To be engaged in learning, students not only require meaningful contexts but also active participation in the knowledge construction process (Hart et al. 2011). According to the learning theory of constructivism, construction of knowledge allows students to develop a deeper understanding of the learning content (Biggs 1999). Authentic pedagogical practices are influenced by the constructivist philosophy of student-centred learning, where students create meaningful knowledge in real-world tasks (Morrissey 2014), thus engaging students in the learning process (Quartuch 2011). How do we ensure that the

authentic tasks require seafarer students to construct knowledge using competencies (technical and soft/underpinning skills) as required in the real-world?

Construct validity evaluates the extent to which the assessment measures the theoretical construct or processes that are internal to an individual (Moskal & Leydens 2000). For example, construct validity ensures that the authentic assessment of a student's ability to fight fires on board a ship not only assesses the technical knowledge of fire-fighting but also the essential and critical underpinning/soft skills of problem solving, communication, and critical thinking. The development of 'soft' skills in students allows them to transfer these skills into different scenarios and roles/responsibilities (Mitchell 2008) and may also create higher student engagement. The recognition of soft skills and the requirement to assess them is essentially missing within the STCW Code. Construct validity can also be achieved through subject experts' validation (Jonsson 2008). Table 3.4 highlights the relevance of construct validity testing to authentic assessment.

Table 3.4: Relevance of construct validity testing to authentic assessment.

Objective	Validity tested	Achieved through
How do we ensure authentic tasks require seafarer students to construct knowledge using competencies as required in the real-world?	Construct validity	Validation by multiple subject experts

Student engagement may be higher if students are provided with clear expectations of learning standards to be achieved before the assessment is implemented (Findlay 2013). Students are then measured against identified standards of achievement. How well the individual student has performed by applying specific criteria and standards (Dunn, Parry & Morgan 2002). Standards are defined as levels of definite attainment and sets of qualities established by authority, custom, or consensus by which student performance is judged, whereas criteria are essential attributes or rules used for judging the completeness and quality of standards (Sadler 2005; Spady 1994). Although, such criterion-referenced assessments are promoted in performance-based assessments like authentic assessment, traditional assessments tend to avoid doing so and follow the norm-referenced assessments (Dikli 2003). Hence, norm-referenced assessments that do not inform students on standards of achievement, if implemented in seafarer education, would not ensure minimum competence (Lister 2006).

Providing students with essential criteria and standards of achievement at the beginning of the learning period is an essential requirement of the authentic assessment process (Wiggins 1989; Archbald 1991; Darling-Hammond & Snyder 2000). In authentic assessment, the teacher provides a roadmap of the entire subject to be learned while

allowing students to construct their understanding of the topic. Providing standards of performance beforehand enables students to reflect on their learning and carry out self-assessments of their thinking and practices towards achievement of the required standards (Findlay 2013). As learning progresses, learners assume increasingly more control over the sequence in which they want to engage their learning (Schell 2000) and gain mastery over knowledge and skills learned through strategic and critical thinking (Fredricks & McColskey 2012). Seafarer students are expected to achieve learning outcomes driven by the STCW Code. However, lack of descriptive outcomes within the code and traditional teaching and assessment practices often do not provide the students with clear expectations of the learning standards to be achieved.

The use of assessment rubrics is one method of providing the students the performance criteria and standards to be achieved in advance (as required in authentic assessment) as well as adhering to the competency standards (Diller & Phelps 2008) such as the STCW code in seafarer education. Rubrics are assessment tools that comprise of individual and essential dimensions of performance known as criteria along with standards for levels of performance against those criteria (Jonsson & Svingby 2007). Using the objective standards and criteria, assessment rubrics can be used for evaluating student performance and providing them with feedback on the level of learning achieved (Diller & Phelps 2008).

Providing feedback on student performance allows educators to identify areas of learning that need improvement. Hence, an assessment rubric can be a very effective tool to obtain inter/intra-rater (scorer or assessor) reliability. Inter-rater reliability evaluates the variations in judgements across raters, while intra-rater reliability examines the consistency of a single rater (Jonsson and Svingby, 2007). The assessment rubrics can be used as a common marking guide by raters, where the objective standards and criteria may promote unbiased marking (Oh et al. 2005). However, to obtain a high inter-rater reliability, rigorous training of raters may be essential to avoid differing approaches to marking (Koh & Luke 2009; Taylor, 2011). Ideally raters should be involved in the development of assessment rubrics, otherwise, it will require time and effort to ensure they understand its purpose and implementation (Diller & Phelps 2008).

On completion of rater marking, assessment rubrics may be used to provide students with feedback on the standards of learning achieved. The feedback may be used by students to engage in meaningful reflection, known as metacognition (Scott 2000). Students reflect on their current level of learning and engage in self-assessment, which allows them to identify the gaps between their current competence and those required by educators or employers in the workplace (Boud & Walker 1998). Recognizing gaps in their knowledge allows students to develop strategies towards filling those gaps making learning more structured and deeper. This is a departure from the ‘surface’ learning

approaches that students engage in purely for passing examinations and hence may engage students in learning. The ability to recognize gaps in knowledge through self-assessment also develops students' understanding of how skills developed in particular contexts may be used in different contexts. This enables seafarer students to understand key requirements for the transfer of learning from the classroom context to ships as a workplace (McCarthy 2013). Table 3.5 highlights the relevance of assessment rubrics as an authentic assessment tool towards enhancing the engagement of seafarer students.

Table 3.5: Relevance of assessment rubrics (authentic assessment tool) towards enhancing student engagement.

Objective	Assessment characteristics	Achieved through
How do we ensure authentic tasks enhance seafarer student engagement?	Clear expectations of learning standards to be achieved provided at the beginning of learning period. Feedback on learning standards achieved provided on completion of learning period.	Assessment rubrics

3.2.2. Outcome: Ability to transfer skills to different contextual scenarios

Students who are able to frequently reflect on their learning to recognize gaps in their own construction of knowledge and improve on them, begin to grasp cues (Leberman 1999; Sator 2000) on applying the same knowledge (developed in a specific context) to different contexts resulting in the transfer of learning (Bransford, Brown & Cocking 2000; Donovan, Bransford & Pellegrino 1999). Authentic assessment implemented as a formative assessment can be used to provide students with different contextual scenarios to apply knowledge gained. Students re-evaluating their learning, develop critical thinking skills causing behavioural changes that promote positive growth in cognitive development, which can be used to assimilate, analyse, and structure information for decision making and problem solving (Saunders et al. 2001). Cognitive development provides students with the belief and confidence (Bandura 1977) to transfer newly acquired knowledge and skills (Merriam & Leahy 2005). Learners draw on and extend previously learned knowledge and develop their own cognitive maps to interconnect facts, concepts and principles. As learning progresses, understanding becomes integrated and structured, leading students to gain mastery over content (Scott 2000). Students' ability to transfer is enhanced when they are able to use the deep understanding of the learning content to interconnect facts and apply it to different contexts (Mestre 2002).

However, according to the constructivism theory of learning, transfer is enhanced when learning is contextualised in authentic tasks designed in meaningful contexts (Ertmer & Newby 1993). Providing authentic tasks that require application of knowledge as in the real-world allows students to identify essential 'threshold concepts' central to facilitating the transfer of learning (Moore 2012). The authentic tasks, which may initially be unfamiliar to students, comprise cues to facilitate understanding of the transfer. The cues allow students to gain an understanding of threshold concepts required to master the subject and to understand how they may be integrated with other units of learning (Cousin 2006). As the complexity of the tasks is increased, fewer cues are provided for students to practice the transfer of learning in dissimilar situations.

Due to the complexity in recreating the shipboard workplace environment in the land-based SET institutes, most of the learning and assessment in seafarer education takes place in decontextualised scenarios. Herrington & Herrington (1998) indicated that authentic assessments conducted in real-world contexts provide 'cues' to students on strategies to adopt when performing in the real world. Contextualised authentic tasks may not recreate all the conditions within a shipboard workplace but may replicate many of the complexities and challenges faced by seafarers in the real-world. Content and construct validity may ensure that the assessment tasks resemble real-world scenarios requiring the targeted competencies to perform adequately within that environment to the required workplace standards. Table 3.6 highlights the relevance of contextual authentic assessment task towards enhancing the ability to transfer skills to different contextual scenarios.

Table 3.6: Relevance of contextual authentic assessment task towards enhancing the ability to transfer skills to different contextual scenarios.

Objective	**Assessment characteristics**	**Achieved through**
How do we ensure authentic tasks enhance ability to transfer skills to different contextual scenarios?	Formative assessment allowing students to self-reflect on learning acquired in real-world task and reapply new learning Task designed in meaningful contexts resembling real-world scenarios	Contextual authentic assessment task

However, capturing a more authentic performance does not ensure validity (Stevens 2013). Testing for internal consistency reliability may be one of the ways to avoid this problem. Internal consistency evaluates how well the different components of the assessment measure a particular construct (Drost 2011). Internal consistency measures 'consistency' within the assessment instrument and based on the average inter-correlations among all the individual items within the test, questions how well the items

measure (Drost, 2011) particular learning outcomes and/or behaviours associated with the learning outcome. Internal consistency reliability can be measured via various statistical measures (Oh et al. 2005; Olfos & Zulantay 2007; Cassidy 2009) and some of these methods (split-half and test-retest reliability) may also generate multiple examples of evidence of competence.

3.2.3. Outcome: Contextual and multiple examples of evidence of competence

Internal consistency reliability can be measured using statistical measures such as Kuder Richardson #20 (Jonsson 2008) or Cronbach's coefficient alpha (Oh et al. 2005), which determine the correlations of the test questions to the competency it purports to measure. This may also be done using the split-half or test-retest reliability (Drost 2011). Split-half reliability involves administering two separate tests or splitting an individual test to create two measures (the results of one half compared with the other) assessing the same construct (Drost 2011). However, irrespective of whether it is a single test or two separate tests, all questions should measure the same construct (McLeod 2013).

Test-retest reliability involves administering the same test after a specific period (Drost 2011). Timing of the test becomes an important variable in this type of reliability test. If the duration between the tests is too short, the students may recall information from their first attempt, which may bias the result. Alternatively, if the duration is too long, there may be a 'learning effect' due to extraneous variables that may not be easily identified (McLeod 2013). In either case of split-half or test-retest reliability, the statistical measures of correlation between test questions provide internal consistency reliability. Additionally, assessing students on two separate tests or the same test twice not only evaluates consistency in performance but also provides multiple examples of evidence of competence and confirms the students' ability to repeat the performance.

Multiple examples of evidence of competence may also be generated if the assessment is tested for criterion validity. Criterion-related validity evaluates the extent to which student scores on an assessment relate to scores on a previously established but valid assessment implemented approximately simultaneously (concurrent validity) or in the future to a measure of some other criterion that is available at a future point in time (predictive validity) (Lang II 2012).

The administration of multiple assessments should also be followed by inter-rater reliability where two or more raters evaluate the student work. The use of assessment rubrics in this case will not only provide evidence of achievement against the learning standards and criteria but also act as a contextual evidence of competence. The rubrics along with the standards and criteria may also detail the context under which the task was performed, and competence acquired. Multiple examples of evidence of

competence may enhance the seafarer employer's perception of the quality of evidence produced via authentic assessment. If the evidence demonstrates the seafarer student's ability to perform authentic tasks that represent real-world scenarios requiring competencies as required at the workplace; and to do so repeatedly and consistently (as verified by multiple raters), seafarer employers may perceive the assessment and the resulting performance to be more valid and reliable. Table 3.7 highlights the relevance of validity and reliability testing towards generating contextual and multiple evidence of competence in authentic assessment.

Table 3.7: Relevance of validity and reliability testing towards generating contextual and multiple evidence of competence in authentic assessment.

Objective	Validity tested	Reliability tested
How do we ensure authentic tasks generate contextual and multiple evidence of competence?	Criterion validity	Test-retest and split-half reliability

3.2.4. Outcome: Valid and reliable student performance

Authentic assessment should not only assess the seafarer students' ability to perform real-world tasks to workplace standards (valid performance) but the ability to do so consistently (reliable performance). Student performance in the assessment tasks should allow valid generalisations about student competence (Wiggins 1992) with respect to the demonstrated learning outcome. However, such generalization cannot be based on one performance, irrespective how complex or authentic the task was (Wiggins 1998). Criterion-related validity or split-half reliability (using two separate tests) of authentic assessments provide students with multiple opportunities to demonstrate mastery over the commonly measured construct in the implemented tests.

Data derived from valid and reliable student performances may be used to identify ways to improve the different aspects of validity and reliability of authentic assessment, which in turn may enhance student performance of tasks. For example, Jonsson (2008) found the overall student scores increased by over 60% when the transparency of rubrics was increased based on student performances in the previous year. This example shows that, although authentic assessment does not ensure enhanced student performance, its validity and reliability testing provides evidence towards change in teaching practices that may result in improved performance.

The previous discussion revealed that authentic assessment can achieve its intended outcomes if it addresses and improves upon the different aspects of its validity and reliability. In the context of seafarer education, if the numerous extraneous variables that affect the validity (content, construct, and criterion) and reliability (inter-rater, internal

consistency, split-half, and test-retest) of the authentic assessment are not improved, then the resulting evidence of competence may become questionable (Olfos & Zulantay 2007) to seafarer employers, adversely affecting the employment of seafarer graduates and defeating one of the key purpose of their education and training. Hence, there was a need for a conceptual and practical framework that has a holistic approach to the validity and reliability of authentic assessment. The required framework was developed in Section 3.3.

3.3. Conceptual framework

The conceptual framework developed in this research identifies and addresses the different aspects of validity and reliability at different stages of the authentic assessment implementation. Based on the definitions of the different aspects of validity and reliability discussed in this paper and their uses in the past research, the development of the framework is discussed in the following three specific stages:

- Before the implementation of authentic assessment;
- During the implementation of authentic assessment; and
- After the implementation of authentic assessment.

3.3.1. Before the implementation of authentic assessment

It is a requirement of authentic assessment to design tasks in a real-world context. Hence, the focus of authentic assessments for validity purposes should be on creating tasks that emulate workplace challenges faced by practicing professionals. Therefore, it is critical that before authentic assessment is implemented, the designed task should be tested against the desired workplace standards to ensure content validity (Moon et al. 2005) and construct validity (Wiggins 1998). Content validity should ascertain whether the authentic tasks resemble real-world scenarios, encompassing wide but required content and assessing only intended outcomes. Thus, content validity is generally attained through a review by subject experts. Construct validity should ascertain whether the task performed required the integration of competence acquired in individual units of learning, using not only technical/occupational skills but also the essential soft/underlying skills. It should also ensure that the tasks comprise forward-looking questions and ill-structured problems as required in authentic assessments. Jonsson (2008) explained that construct validity can also be achieved through subject experts' validation before the authentic assessment is implemented.

The performance criteria should be provided beforehand, at the beginning of the learning period to the students. This should preferably be carried out through assessment rubrics as they detail the essential criteria and standards to be achieved by

the students. Providing assessment rubrics beforehand allows the students to use them as a guide before and during assessments to develop strategies towards the collection of evidence required to demonstrate competence at the required standards of learning.

3.3.2. During the implementation of authentic assessment

Once the authentic assessment is implemented, the student performance should be marked using the inter-rater reliability approach. The inter-rater approach uses more than one rater (scorer) to ascertain the consistency of the results. The assessment rubric is useful to the scorers as it provides them with clear guidelines on the essential criteria and standards of performance expected from the students. Using the same rubric for assessment and marking ensures objectivity and fairness in the results. In evaluating scores involving raters, it is important to know the extent to which different scorers agree or disagree on the values assigned to student responses (Moon et al. 2005). Cases where multiple raters do not agree with the values assigned to student performance may produce a discrepancy in the resulting evidence of competence and create employer dissatisfaction. Hence, to establish more consistency and reliability in scoring, the framework may need to adopt a practical approach of using a two-member rater panel with a third panel member available for arbitration in case of a disagreement between the raters (Taylor 2011).

3.3.3. After the implementation of authentic assessment

Once authentic assessment is implemented, and the initial evidence of competence is acquired, the framework should establish internal consistency reliability to determine the degree to which individual items that comprised the assessment, consistently measure the same objectives. Finally, the framework should employ criterion validity to compare the effectiveness of the authentic assessment task to measure the professional competence with a secondary assessment. The secondary assessment should be an existing but valid assessment that measures the same construct (Gao & Grisham-Brown 2011) and may be implemented concurrently or at a later date. Employing concurrent validity generates multiple examples of evidence of competence to perform the task and the students' ability to use the underlying competencies.

The effectiveness of the framework to address the validity and reliability of authentic assessment and of its ability to generate the stipulated outcomes is verified via a feedback loop provided at the end of the framework. This is because the effectiveness of the valid and reliable authentic assessment of students is ascertained only after the event. Data from student performances provide valuable inputs towards the improvement of the assessments. While student and employer perceptions provide

feedback on the authentic assessment outcomes, data from student performances provide the necessary feedback to enhance the validity and reliability of authentic assessments. Once the feedback is obtained, the loop takes the educators back to the design stage of the assessment task. Modifications based on the feedback enhance the validity and reliability of authentic assessments; and in turn the resulting outcomes of assessment.

Figure 3.1 describes the conceptual framework created to address the validity and reliability of authentic assessment when implemented for SET. The authentic assessment framework for SET (AAFSET) employs a holistic approach to the validity and reliability of the authentic assessment. However, the framework is conceptual in nature and must be tested.

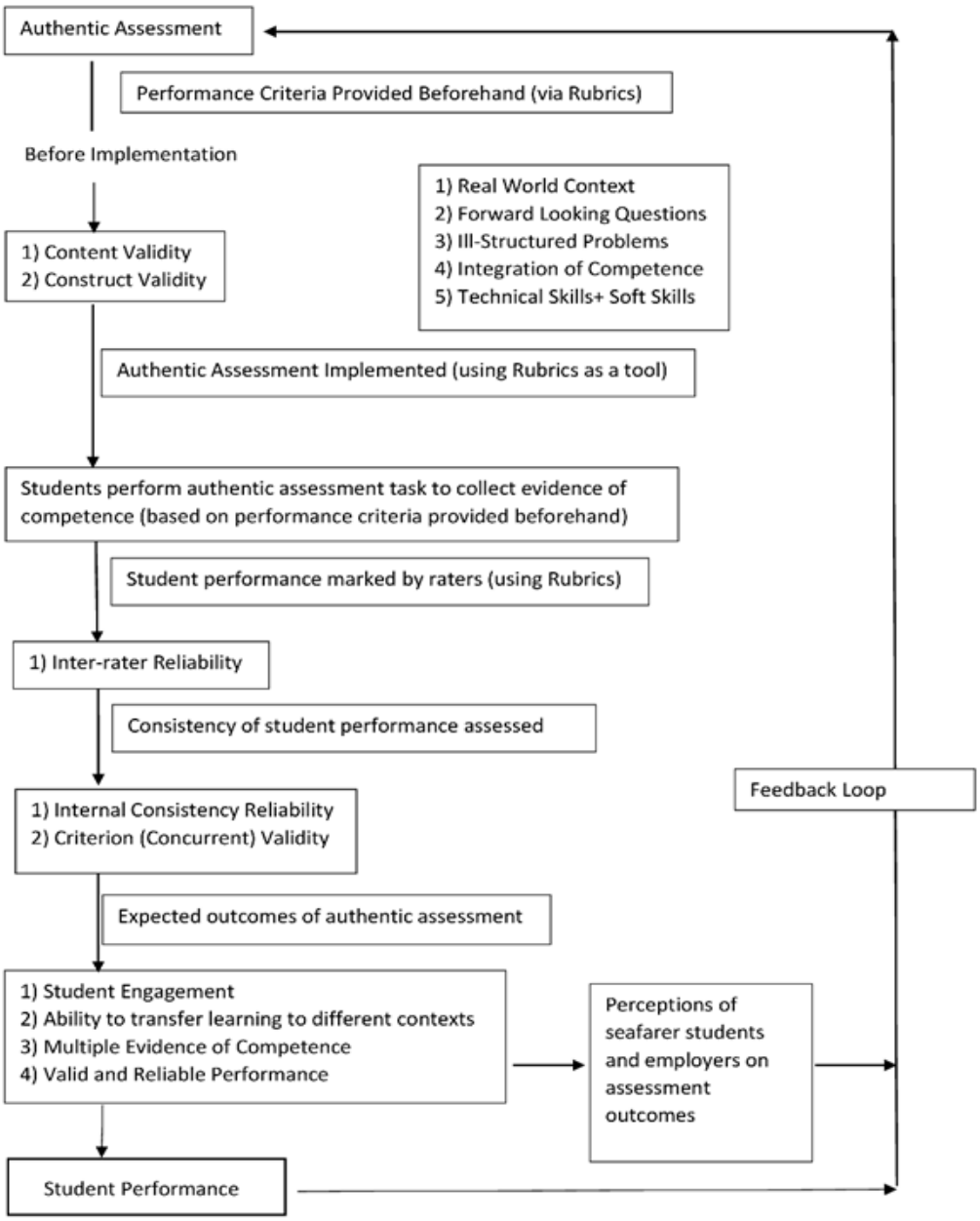


Figure 3.1: Authentic assessment framework for seafarer education and training (AAFSET).

The next chapter (Chapter 4) of the thesis will focus on the research methodology used in this study. In particular, it will explain how the quantitative data collection and analysis techniques were used to answer the research questions formulated in Chapter 1.

4. Methodology

The purpose of this chapter is to describe the research methodology employed to investigate the research questions. The following research question (RQ)s directed the data collection and analysis; and served as a framework for the interpretation of the data:

RQ1: Is there a significant improvement in seafarer students' academic achievement in authentic assessment when the scores are compared with traditional assessment scores?

RQ2: Is there a significant improvement in seafarer students' academic achievement in the formative authentic assessment when the scores are compared with summative traditional assessment scores?

RQ3: What is the correlation between seafarer students' perception of authenticity in assessment and their academic achievement in the associated assessment tasks?

This chapter is written in two parts. Part 1 investigated RQs 1 and 2; and part 2 investigated RQ3. In terms of publications included in this thesis, RQs 1 and 2 were investigated in Paper VI, and RQ3 was investigated in Paper VII.

4.1. Research methodology – Part 1

The objective of this research project was to investigate if authentic assessment increased seafarer students' academic achievement (through the comparison of scores obtained) as compared with traditional assessments. Separate and independent seafarer student groups were identified as the control (traditional assessment) and the treatment group (authentic assessment). The traditional and authentic assessments were implemented in the selected unit of 'Managerial and leadership skills' within the Bachelor of Nautical Science degree program at the Australian Maritime College (AMC), an institution of the University of Tasmania (UTAS). The Bachelor program of study is provided for students who intend to embark on a career in the maritime industry as ranked officers on commercial ships. It provides the knowledge and skills required to safely manage and operate ships. The unit of 'Managerial and leadership skills' was selected since it enrolled the highest number of students within the degree program. Higher number of students maximized the participants and hence, enhanced the robustness of findings.

The authentic assessment implemented for the selected unit differed from the decontextualised traditional assessment on the basis of the inclusion of a real-world context that attempted to closely replicate the complexities and challenges faced by seafarer students on ships through a simulation of the scenarios described in the case

studies used for both types of assessments. The inclusion of the real-world context being the only differing aspect between the two types of assessments, the 'authenticity' (provided through a real-world context) of the assessment was the focus variable.

However, according to past researchers (Bailey 1998, p. 205; Law & Eckes 1995; Dikli 2003, p. 16; Abeywickrama 2012) traditional assessments have been conventionally described as not only inauthentic but also as a "one-shot" and single-occasion tests implemented at the end of learning (summative) period. Hence, the scores obtained in the summative traditional assessments cannot inform on the progression of the learner as they only measure the students' ability at a particular time (Law & Eckes 1995). This is also the case with the oral examinations conducted to assess the seafarer's competence before issuing them with the CoC. Seafarers who are unable to answer the questions to the satisfaction of the assessor are declared as 'fail' before being provided with another opportunity which often demoralises the students (Prasad 2011).

In comparison to the summative traditional assessments, one of the key characteristics of authentic assessment, as defined by its major authors (Wiggins 1989; Archbald 1991; Gulikers 2006), required students to be informed on their gaps in knowledge through feedback on their first attempt at the assessment task; and then provided with at least one opportunity (formative) to improve their performance in a similar task at a different time (Law & Eckes 1995) before making the final judgement on their competence.

Hence, the authentic assessment in this research project was implemented as a formative assessment and in comparison, the traditional assessment was summative in nature. The objective of distinguishing the two assessments based on their implementation was to collect valuable empirical evidence that would justify either the continuation or the change in summative assessment methods currently used in seafarer education. Since the 'nature of task implementation' (formative versus summative) was a differing aspect between the two types of assessment, an additional variable (apart from 'authenticity') was introduced in this research. Hence, this research also investigated the difference in seafarer students' academic achievement comparing scores of the formative authentic assessment with the summative traditional assessment. Due to the nature of the assessment tasks (students were required to respond to questions based on a case study), additional independent variables (work experience, English as the first language, and educational qualification) based on their ability to influence student performance and the resulting academic achievement were also identified. The student scores were isolated on the basis of the independent variables and analysed to investigate the effect of these variables on students' academic achievement. The findings of this research project revealed recommendations for education and training providers towards the implementation of authentic assessment and improvement of students' academic achievement.

4.1.1. Research design

The difference in seafarer students' academic achievement (traditional versus authentic) for the unit of 'Managerial and Leadership Skills' was investigated in this research project. Students completing this unit acquire the knowledge and skills required by a senior seafarer officer to organise and manage the efficient operation onboard a merchant ship. The unit focuses on leadership and management of multicultural crews in a global environment and the maintenance of an effective interface with other industry stakeholders.

The students that enrolled in the unit in semester 1 were classified as the 'control group' that underwent a traditional assessment. The traditional assessment comprised of two case study scenarios (as shown in Appendix 10) presented and described only on paper in absence of a real-world context. The students provided written responses on paper to essay-type questions based on their analysis of the described scenarios relying solely on their ability to recall how the scenarios would have played out in the real-world onboard ships.

In comparison, another cohort of students enrolled in the same unit in semester 2 were assessed authentically through the same case studies (as shown in Appendix 10) described on paper. Although, the authentically assessed students also provided written responses on paper to the same essay-type questions, the authentic assessment differed from the traditional assessment by providing a real-world authentic context to the assessment task through a simulation and practical demonstration of the same case study scenarios, as employed in the traditional assessment, enacted by AMC staff. For example, one case study that described ship staff abandoning the ship using a liferaft during a fire was demonstrated at AMC training pool. The pool was equipped with facilities to launch a real liferaft in simulated waves, strong winds, darkness, rain, and smoke. The simulation also included ringing of the emergency alarms and staff playing the role of panicking seafarers jumping into the pool to replicate a possible emergency. In comparison to the authentic assessment, students assessed traditionally relied only on their imagination and experience to visualise the described scenarios.

Although one may argue that the descriptive case studies in itself (without the simulation) may have provided the real-world contexts, the simulations engaged the sensory perceptions of the students requiring them to demonstrate the ability to analyse, assimilate, and integrate presented information and construct responses towards it. This was similar to the workplace where professional seafarers analyse available information and take required action, and thus, distinguished the traditional from the authentic assessment.

In addition to the authentic design, the assessments also differed in the nature of their implementation. The authentic assessments were formative in nature and held on two different days (3 weeks apart). The second authentic task was implemented once the students received individual feedback on their performance in the first authentic task. In comparison to the authentic assessment, the traditional assessment was summative in nature and both case studies were implemented at the assessment. However, the duration of the authentic assessment (combined) was the same as that of the traditional assessment. The assessment details and rubric were provided to both the student groups at the beginning of the semester. To avoid the introduction of additional variables, the unit, learning content, lecture delivery methods, lecturer, assessment rubric, total duration of the assessment, and assessment questions were kept constant. The number of completed semesters and academic workloads were the same for both groups. Both the assessments were supervised by external invigilators appointed by AMC. Table 4.1 summarises the research design.

Table 4.1: Summary of research design.

Unit of competence	Managerial and leadership skills	Managerial and leadership skills
Participants	Seafarer students enrolled in the Bachelor of Nautical Science degree program	Seafarer students enrolled in the Bachelor of Nautical Science degree program
Group	Control Group	Treatment group
Semester	1	2
Assessment type	Traditional assessment	Authentic assessment
Sample size	96 students	93 students
Task description	Students respond to case studies described in the assessment	Students are provided with a real-world 'authentic' context for the case study described in the assessment
Nature of assessment	Summative	Formative
Task implementation	Two case studies implemented together	One case study implemented three weeks apart (Total: 2 case studies)
Response method	Written response to essay-type questions	Written response to essay-type questions
Duration	One hour	30 minutes for each case study (Total: 1 hour)

4.1.2 Data analysis

The quantitative data (assessment scores) was analysed using MS Excel. The student scores were analysed using the values of mean scores, standard deviation, effect size, and the t-test values. While the mean scores provided an indication on the difference in students' academic achievement between the two types of assessments implemented, standard deviation informed on the scattering of the individual scores in each type of

assessment to indicate the variation. The recommended (Coe 2002) effect size (0.5 or greater) and the t-test values (P <.05) indicated if the variation in scores of students' academic achievement was statistically significant for reporting.

4.1.3 Sampling considerations

The sampling technique used in this research was based on convenience sampling that relies on opportunity and participant accessibility, and used when the study population is large, and the research is unable to test every individual (Clark 2014; Robson 2011). Participants for this research were seafarer students drawn from the course of Bachelor of Nautical Science enrolled in the selected unit at AMC. This research was based on the sample of 96 participants (as the control group), and 93 participants (as the treatment group). Scores of seven students from the control group and nine students from the treatment group were not included in the analysis due to the failure of the students to complete the administrative paperwork. A key consideration while sampling was to ensure that the control and treatment groups comprised of randomly assigned students where each participant had an equal chance of participating in this research based only on the sequence of enrolment in the individual semesters. The groups were not sorted based on any other pre-determined characteristics, such as qualifications, academic ability or work experience that may have impacted the outcomes of this research.

4.1.4 Validity and reliability of assessment

4.1.4.1 Before implementing assessment

Content and construct validity was achieved by using a jury of experts before the assessment was implemented. The subject experts comprised of seven field experts (leadership and teamwork skills in a maritime context) within AMC. The subject experts included ex-seafarers currently employed as educators in the field of seafarer education, each having more than 25 years of work experience in the maritime industry. The first draft of the assessment instrument was sent to the subject experts who were asked to make recommendations towards improving the instruments. The experts provided suggestions on simplifying terms used in the case study for universal understanding. For example, the words 'imperative', 'mitigate', and 'hinder' were substituted with the words 'vital importance', 'reduce', and 'delay'. Suggestions were also provided on the distribution of marks, length of the tasks, and ways to demonstrate the case studies authentically within the educational settings at AMC. The details of the experts are provided in Table 4.2.

Table 4.2: Details of subject experts.

Expert number	Qualification (Professional Mariner)	Highest Academic Qualification	Work Experience (Professional Mariner)	Work Experience (Maritime Educator)
1	Master Mariner	PhD	16	21
2	Master Mariner	Bachelor	17	24
3	Master Mariner	PhD	18	15
4	Master Mariner	Master	22	13
5	Chief Mate	Bachelor	12	23
6	Chief Engineer	Bachelor	20	24
7	Chief Engineer	Master	25	11

4.1.4.2 After implementing assessment

Criterion validity for authentic assessment was obtained with a secondary authentic assessment implemented three weeks after the first assessment. The test for criterion validity allowed to assess the consistency of student performance in authentic assessments.

To establish more consistency, objectivity, and reliability, the student scores were reviewed by the panel of the subject experts using the assessment rubric.

4.1.5 Ethics approval

A minimal risk ethics application approval, constituting ethical clearance by the Tasmania Social Sciences Human Research Ethics Committee, was obtained for this research project. The ethics approval is attached as Appendix 9. Participants were reassured that the data would be anonymised and that their contribution would be confidential. Students were free to withdraw at any time from the study.

4.1.6 Research questions, hypothesis, and independent variables

The focus of this research project was to investigate the difference in seafarer students' academic achievement by comparing traditional assessment scores with authentic assessment scores. Hence the following research questions (RQ) were developed:

RQ1: Is there a significant improvement in seafarer students' academic achievement in authentic assessment when the scores are compared with traditional assessment scores?

RQ1 enabled the development of the following research hypothesis (H_{1a}):

1a) H_{1a}: There is a significant improvement in seafarer students' academic achievement in authentic assessment when the scores (AA) are compared with traditional assessment scores (TA). This hypothesis was denoted as: *Score AA > Score TA*

As stated in the 'Research design' section, the authentic assessment was implemented as two separate tasks (or case studies) with the second task implemented three weeks after the first task. The authentically assessed students received individual feedback on their performance in the first task before attempting the second authentic task three weeks later. The feedback to students was provided individually using the assessment rubric (attached as Appendix 8) that defined the standards and criteria of performance achieved by the students. Additional to the rubric, feedback comments were also included in the students' answer sheets for their perusal. Finally, a generalised feedback was also provided to the students as a group in the classroom and using the online learning tool. In comparison, the traditional assessment implemented both tasks at the same assessment. Hence, the traditionally assessed students did not receive individual feedback on the first task to improve their performance in the second task. Since, the first task in both traditional and authentic assessments were performed without any prior feedback, and the differing aspect between the assessments was only the 'authentic' nature, the next hypothesis was also developed towards answering RQ1.

1b) H_{1b}: There is a significant improvement in seafarer students' academic achievement in authentic assessment when the scores for the first task (AA_1) are compared with traditional assessment scores for the first task (TA_1). This hypothesis was denoted as: *Score* AA_1 *>* *Score* TA_1

It was evident from the differing nature of assessment implementation (formative versus summative) that contrary to students assessed authentically, students assessed traditionally did not receive an opportunity to improve their academic achievement based on feedback. Thus, apart from the 'authentic' design, additional variables (an opportunity to improve achievement in authentic assessment) that may have influenced student achievement in this research were introduced due to the nature of assessment implementation. Hence, the difference in seafarer student achievement by comparing scores obtained in summative traditional assessment with scores obtained in formative authentic assessment was investigated in this research project. This resulted in the development of the following RQ:

RQ2: Is there a significant improvement in seafarer students' academic achievement in the formative authentic assessment when the scores are compared with summative traditional assessment scores?

RQ2 enabled the development of the following research hypothesis:

2a) H_{2a}: There is a significant improvement in seafarer students' academic achievement in authentic assessment when the scores for the second task (AA_2) are compared with traditional assessment scores for the second task (TA_2). This hypothesis was denoted as: *Score AA_2 > Score TA_2*

To answer RQ2, it was necessary to investigate the difference in the students' academic achievement if the assessment design was kept constant, and the only differing aspect between the student performances was the nature of assessment implementation. It was assumed that authentically assessed students that received feedback on their performance in the first task and an opportunity to improve on their performance, would achieve higher scores in the second task. Hence, keeping the 'authentic' design of the assessment as a constant, the following hypothesis was developed:

2b) H_{2b}: There is a significant improvement in seafarer students' academic achievement in authentic assessment when the scores for the second task (AA_2) are compared with the scores for the first task (AA_1). This hypothesis was denoted as: *Score AA_2 > Score AA_1*

Since, the summative nature of the traditional assessment did not allow students to receive individual feedback on their performance in the first task to recognize gaps in their knowledge; and another opportunity to improve their academic achievement in the second task, it was assumed that traditionally assessed students would find it challenging to significantly improve their academic achievement in the second task. Hence, keeping the 'traditional' design of the assessment as a constant, the following hypothesis was developed:

2c) H_{2c}: There is no significant improvement in seafarer students' academic achievement in the second traditional assessment task (TA_2) when the scores are compared with the scores for the first task (TA_1). This hypothesis was denoted as: *Score TA_2 ~ Score TA_1*

The research questions and the resulting hypotheses is summarised in Table 4.3.

Table 4.3: Research questions and the resulting hypothesis.

Research Question	**Hypothesis**
RQ1: Is there a significant improvement in seafarer students' academic achievement in authentic assessment when its scores are compared with traditional assessment scores?	*H_{1a}: Score AA > Score TA* *H_{1b}: Score AA_1 > Score TA_1*
RQ2: Is there a significant improvement in seafarer students' academic achievement in formative authentic assessment when its scores are compared with summative traditional assessment scores?	*H_{2a}: Score AA_2 > Score AA_1* *H_{2b}: Score AA_2 > Score TA_2* *H_{2c}: Score TA_2 ~ Score TA_1*

4.1.7 Independent variables

This research focussed on investigating the difference in students' academic achievement that may have resulted either due to the design of the assessment (traditional versus authentic) or the nature of its implementation (summative versus formative). Hence, this research required to identify the independent variables that could influence the achievement. Since, both traditional and authentic assessments implemented in this research required seafarer students to respond to case study scenarios, the independent variables identified were based on their efficacy with regards to influencing student performance and resulting scores. Thus, the following variables were identified:

- **Work experience**: The assessment tasks required students to respond to case study scenarios based on situations that they might encounter on board ships. There was a possibility that students with higher work experience may have encountered similar situations and hence, were better equipped to answer the questions. Although it was not a stringent requirement, students enrolled in the selected unit were expected to have completed the minimum work experience of one and half to three years on ships. Thus, the extraneous variable of 'work experience' was classified as students with 'less than three years' and 'more than three years' of experience.

- **English as first language**: Since students were required to provide written responses describing their actions in the case study scenarios, proficiency in the English language could significantly affect their ability to provide descriptive answers. This research project does not imply that all non-native English speakers do not have proficiency over the language. Since, this project did not conduct any additional tests to assess the English language proficiency of non-native English speakers, it was necessary to distinguish them from students with English as their first language.

- **Level of education completed**: The minimum requirement for enrolment in the bachelor's program is a senior secondary school (Grade 10 - Grade 12) qualification. However, the selected sample for this research included students with qualifications higher than Grade 12 including those with under-graduate or post-graduate qualification from universities. Students completing higher academic qualifications such as university studies may be better equipped in their ability to analyse and respond to case study scenarios compared with students who have only completed studies at school level. Hence, the variable of 'level of education completed' was classified as students who had completed up to high school (Grade 10 - 12) and students who had completed education higher than Grade 12.

4.2. Research methodology – Part 2

Investigations of RQs 1 and 2 confirmed that seafarer students' academic achievement was significantly higher in the formative authentic assessment when compared with the summative traditional assessment (discussed in Chapter 5). Although, in past research work, higher academic achievement was attributed to the 'authentic' design of the assessment and the formative nature of its implementation, further research was required to investigate the factors of assessment that the students may have perceived significant and influenced their perception of authenticity in assessment leading to higher academic achievement. These factors will provide guidance to assessors in the designed authentic assessment with the aim of improving scores and the resulting academic achievement. Hence, using the same but independent sample of authentically assessed students, the research presented in this paper investigated student perceptions of authenticity in assessment to reveal the factors of assessment that correlated significantly with their academic achievement.

As a result, RQ3 was developed:
RQ3: What is the correlation between seafarer students' perception of authenticity in assessment and their academic achievement in the associated assessment tasks?

The developed RQ enabled the development of the following research variables:

- independent variable: Perceptions of authenticity in assessment; and
- dependent variable: Students' academic achievement.

This research identified seafarer students' 'perception of authenticity in assessment' as the independent variable. The term 'authenticity' in this regard referred to the characteristics (e.g. setting assessment tasks in real-world contexts) of the authentic assessment that students may perceive significant towards the outcomes of: higher student engagement; ability to transfer skills to different contexts; contextual and multiple examples of evidence of competence; and valid (relevant to workplace) and reliable (multiple and consistent) student performance. The defining characteristics of authentic assessment that lead to the aforementioned outcomes are explained in Chapter 3 of this thesis; and summarised in Table 1. Subsequently, the key words (bold in Table 4.3) in the defining characteristics of authentic assessment were used to conceptually develop the factors of assessment (task, context, criteria, etc.). The development of the factors is also shown in Table 4.4.

Table 4.4: Defining independent variable to provide conceptually developed factors of assessment for measuring seafarer students' perception of authenticity.

Independent variable	**Defining characteristics**	**Conceptually developed factors of assessment derived from keywords in the defining characteristics**
Perception of 'authenticity' in assessment	Assessment outcomes: Higher student engagement; Ability to transfer skills to different contexts; Contextual and multiple evidence of competence; Valid and reliable student performance	
Authentic assessment outcomes:		
Higher student engagement	Setting assessment **tasks** in real-world **contexts**; Assessment tasks should be **relevant to the workplace;** Assessment's emphasis on active **construction of knowledge**; Performance **criteria** should reflect workplace needs and be provided beforehand to show **transparency**; **Multiple opportunities** for students to improve learning based on **feedback** on learning achieved	Task; Context; Relevance to the workplace; Construction of knowledge; Criteria; Transparency of criteria; Multiple opportunity based on feedback
Ability to transfer skills to different contexts	Setting assessment **tasks** in real-world **context**; Students using **feedbac**k to identify and fill gaps in competence through **multiple opportunities**	Task; Context; Multiple opportunity based on feedback
Contextual and multiple evidence of competence	Students provided with **multiple opportunities** to improve learning based on previous **feedback**	Multiple opportunity based on feedback
Valid and reliable student performance	Assessment tasks should be **relevant to workplace**; **Multiple opportunities** to improve learning based on previous feedback	Relevance to the workplace; Multiple opportunity based on feedback

Based on the conceptually developed factors (Table 4.4), this project adapted a questionnaire [majorly from Gulikers (2006)] which was used to obtain student responses regarding their perception of authenticity in assessment. In Stage 1, the perceptions of authenticity for the conceptually developed factors were correlated to the dependent variable of students' academic achievement (defined by their composite numeric scores obtained in the authentic assessment tasks). Stage 2 extracted new factors of assessment through a factor analysis. Using the student responses from the perception survey, an additional correlational analysis was conducted between students'

perception of authenticity for the new factors of assessment and their scores in the authentic assessment.

Both stages of investigation revealed significant findings towards the design of authentic assessments for higher academic achievement of students.

4.2.1. Questionnaire design

This paper used a questionnaire to measure seafarer students' perception of authenticity in assessment. To develop the questionnaire, past research in the area of authentic assessment was scanned to investigate if existing published questionnaires and/or items could be used for the purpose. Additionally, an internet search was conducted for the same purpose. The final survey document (attached as Appendix 4) developed for this research used all the questions from Gulikers (2006) to form Questions 5—27. Since Guliker's (2006) questionnaire was developed for social work students, the word 'social worker' was replaced with the word 'seafarer' in the questionnaire developed for this project. One question was adopted from the National Survey of Student Engagement (NSSE) to form Question 28a—28e. Two questions were devised by the authors of this paper to form Questions 29 and 30a—30b. The first four questions enquired student demographic details. Questions 5—27 and 29—30 were scored on a 5-point Likert scale ranging from 1 (totally disagree) to 5 (totally agree). Only Question 28 was scored on a 4-point Likert scale ranging from 1 (very little) to 4 (very much). The Likert scale was reverse coded for negatively worded questions (i.e. Questions 10, 11, 18, 23, 26, 28a). Question 30a required a response on the nominal scale of 'Yes' or 'No'.

4.2.2 Validity and reliability of the questionnaire

Since the questionnaire constructed for this research was mainly drawn (barring three questions) from Gulikers (2006), it initially derived its validity and reliability from the values published by that author. According to Gulikers (2006), all scales of the survey had a reasonable internal consistency, shown in Cronbach's alpha (α) ranging from 0.63 to 0.83. The α for the survey used in this research had a value ranging from 0.69 to 0.75. The adaptation of Guliker's questionnaire for the purposes of this research study was validated through an expert validation process. The questionnaire was reviewed through a pilot survey by 12 fellow academics and researchers within the AMC, where the research was conducted. The pilot survey respondents suggested retaining most of the original questions but defining the terms 'context', 'criteria', 'oriented', 'under-graduate', post-graduate', and 'output' used in the survey, for the students. The respondents also suggested excluding the demographic question enquiring the age of

the students and including the question related to the educational qualifications. The details of the pilot survey respondents are provided in Table 4.5.

Table 4.5: Details of the pilot survey respondents.

Respondent number	Position	Highest Academic Qualification	Work Experience (Educator)	Department
1	Professor	PhD	21	Seafaring
2	Professor	PhD	27	Maritime Logistics
3	Senior Lecturer	PhD	15	Seafaring
4	Lecturer	PhD	7	Marine Engineering
5	Lecturer	PhD	8	Maritime Logistics
6	Senior Lecturer	PhD	10	Marine Engineering
7	Lecturer	PhD	4	Maritime Logistics
8	Lecturer	Master	28	Seafaring
9	Lecturer	Master	11	Seafaring
10	Lecturer	Bachelor	23	Seafaring
11	Senior Lecturer	PhD	5	Seafaring
12	Lecturer	Master	24	Seafaring

4.2.3. Data collection

The survey was administered on completion of the authentic assessments for the treatment group. A general announcement was made in class and an email was sent inviting students to participate in the survey. A minimal risk ethics application approval, constituting ethical clearance by the Tasmania Social Sciences Human Research Ethics Committee, was obtained for this research project. The ethics approval letter is attached as Appendix 9. Participants were reassured that the data would be anonymised and that their contribution would be confidential. Students were free to withdraw at any time from the study.

4.2.4 Sampling considerations and response rate

The sampling technique used in this research was based on convenience sampling that relies on opportunity and participant accessibility and is used when the study population is large, and the research is unable to test every individual (Clark 2014; Robson 2011). Thus, participants for this research were two separate groups of seafarer students drawn from the Bachelor of Nautical Science programme at AMC enrolled in the selected unit. A key consideration while sampling was to ensure that the treatment group was comprised of randomly assigned students in which each participant had an equal chance of being chosen based only on the sequence of enrolment in the individual semesters. The groups were not sorted based on any other pre-determined characteristics, such as

qualifications, academic ability, age or work experience that may have impacted the outcomes of this research. This ensured that the relationship between the two variables remained the same in all segments of the sample, which is essential for correlational research (Graziano & Raulin 2000). Moreover, in correlational research the coefficient of determination (r^2) that allows us to estimate how useful the relationship between the dependent and independent variables might be in a prediction (and is a measure of effect size), should be considered significant only if the minimum sample size is 30 (Suresh & Chandrashekara 2012; Lodico, Spaulding & Voegtle 2010; Blondy 2007; Graziano & Raulin 2000). This research, thus, exceeded the recommended minimum sample size.

Although 102 students were asked to respond to the survey, only 98 students participated in the study. Out of the 98 respondents, only 93 surveys were usable for analysis, as 5 surveys were discarded due to incomplete/absent responses.

4.2.5 Data analysis

The correlation analysis was conducted in two stages using the statistical software package SPSS 23.

4.2.5.1 Stage 1: Correlation analysis between students' perception of authenticity in assessment (for factors derived conceptually) and their scores

The questionnaire statements were categorised under the conceptually developed factors of assessment (task, context, criteria, etc.) as determined in Table 4.4. Questions categorised under a common factor were subjected to an inter-reliability analysis (α) to ensure that they were significantly correlated to each other. This is detailed in Table 4.6.

Table 4.6: Survey questions categorised under conceptually developed factors of assessment; and their inter-reliability values.

Question number	Question statement	Factors of assessment	Cronbach's alpha
5	This assessment was oriented to my future profession of a seafarer.	Relevance to the workplace	0.840
6	This assessment was clearly directed to my professional requirements.		
7	This assessment prepared me for my future profession.		
15	This way of assessing is an effective way of assessing professional skills.		
16	This way of assessing fits well with the seafarer's profession.		
17	The output that I had to produce in this assessment is part of the seafarer's job.		
18	The output that was evaluated in this assessment is different from what is being evaluated in practice.		
19	The result that I had to produce in this assessment is something that a real seafarer also has to produce in practice.		
27	In this assessment, both knowledge and professional skills were important.		
8	The task of the assessment resembled the task of a real seafarer.	Task	0.478
9	The task of this assessment was an important part of the seafarer profession.		
10	The task of this assessment differed from the tasks of a real seafarer.		
11	The context in which I had to perform the assessment was fake.	Context	0.650
12	The context in which I had to perform the assessment looked like a seafarer's workplace.		
13	The context in which I had to perform the assessment looked just like the real world.		
14	The context in which I had to perform the assessment was realistic.		
20	The criteria resembled the criteria that I have to meet in practice.	Criteria	0.547
21	The criteria that I had to meet in this assessment resembled the criteria used in practice.		
22	In this assessment, I was evaluated on criteria important for the seafarer's profession.		
23	In this assessment, I was evaluated on things that I never have to use in real profession practice.		
24	The criteria that I had to meet in this assessment were clear enough.	Transparency of criteria	0.763
25	Before I started the assessment, it was clear to me what was expected of me.		
26	It was hard to find out what was expected of me in this assessment.		
28	The following requirements of the assessment helped me to improve my score:	Construction of knowledge	0.540
28a	Memorising course material		
28b	Applying facts, theories, or methods to practical problems or new situations		
28c	Analysing an idea, experience, or line of reasoning in depth by examining its parts		
28d	Evaluating a point of view, decision, or information source		
28e	Forming a new idea or understanding from various pieces of information		
29	The feedback provided in this assessment helped me to identify the strengths and weaknesses in my learning.	Multiple opportunity based on feedback	0.697
30a	This assessment provided more than one (1) opportunity to improve my score.		
30b	If 'Yes', the feedback provided on my first performance helped me to improve my assessment score in the next performance.		

For the purposes of this paper, a α value of greater than 0.70 (Tavakol & Dennick 2011) was considered statistically significant for reporting. Table 4.4 showed that an inter-reliability analysis of the categorised survey questions revealed an acceptable α value (0.70 or greater) for only two factors of assessment, i.e. relevance to workplace and transparency of criteria. Since, an acceptable value of α was found for a low number, i.e. only two factors, a correlation analysis between seafarer students' perception of authenticity in authentic assessment for all the developed factors and their scores in the associated assessment task was conducted. The correlation between the variables (perception of authenticity and scores) was considered significant if the correlation coefficient (R) value was higher than 0.25 (Clark 2014). The findings of the correlation analysis conducted in stage 1 are discussed in the 'Results' section later on in this paper.

4.2.5.2 Stage 2: Correlation analysis between students' perception of authenticity in assessment (for factors extracted through factor analysis) and their scores

Since, the majority of the conceptually developed factors of assessment (except transparency of criteria and relevance to workplace) had a low value (less than 0.70) of α, a factor analysis to statistically develop new factors of assessment was conducted. Next, a factor analysis to remove multicollinearity and extract factors that are relatively independent of one another was conducted. Factors extracted after the exploratory factor analysis (EFA) and their contribution towards explaining the variance in data are shown in Appendix 5.

Appendix 6 reports the loading of the survey questions under the factors derived from the factor analysis. The questions loaded cleanly (without overlap) under the seven factors. The construction of knowledge questions (28b—28d) clustered in Factor 2, the 'context' questions (question 12—14) item in Factor 4, the 'transparency of criteria' questions (24—26) in Factor 5, and the 'multiple opportunity' questions (29—30b) in Factor 7. Hence, these factors retained the original titles. The questions that were reverse coded clustered in Factor 6, which was therefore titled irrelevant to the profession.

Conversely, the questions related to the conceptually developed factors of relevance to the workplace, task and criteria did not cluster in the expected way; and loaded unevenly (split loading) in Factors 1 and 3. Although a limitation of factor analysis is that factor names may not accurately reflect the variables within the factor, especially in the case of split loadings (Yong and Pearce, 2013), this research used the factor naming technique suggested by Neill (2008). Neill advocated for using the majority of the loading items for naming each factor. The items in Factors 1 and 3 were reviewed to provide meaningful names for the extracted factors based on the top loadings for each factor. Additionally, each factor was subjected to an inter-reliability analysis (α) to verify if the values were greater than 0.70. Table 4.7 details the survey question numbers with their factor

loadings, together with the factor titles, and the Cronbach's value of inter-reliability analysis.

Table 4.7: Factors extracted using factor analysis: Categorised survey questions, titles, and inter-reliability values.

Factor	**Survey questions**	**Factor title**	**Cronbach's alpha**
Factor 1	5, 6, 7, 8, 9, 17, 22	Relevance to the profession	0.865
Factor 2	28b, 28c, 28d, 28e	Construction of knowledge	0.806
Factor 3	15, 16, 19, 20, 21, 27	Assessing competence to job-relevant criteria	0.868
Factor 4	12, 13, 14	Context	0.732
Factor 5	24, 25, 26	Transparency of criteria	0.763
Factor 6	10, 11, 23	Irrelevant to the profession	0.616
Factor 7	29, 30b	Multiple opportunity	0.697

Based on the inter-reliability values of α, Table 4.7 revealed that the factor analysis extracted five factors with an acceptable value of more than 0.70. Factors 6 and 7 were rejected due to low α values of less than 0.70. The selected factors (1—5) cumulatively explained 60% of the variance in the data, which was considered significant (Williams, Brown & Onsman 2010) for further correlation and regression analysis. Thus, stage 2 investigated the correlation between seafarer students' perceptions of authenticity for the new factors (1—5) of assessment extracted through factor analysis and their scores in the associated assessment task. The correlation between the variables (perception of authenticity and scores) was considered significant if the correlation coefficient (R) value was higher than 0.25 (Clark, 2014). The findings of the correlation analysis conducted in stage 2 are discussed in the 'Results' chapter (Chapter 5).

The results of the quantitative analysis will be presented in the next chapter (Chapter 5).

This chapter details the findings obtained from performing the quantitative analysis detailed in Chapter 4. Tables 5.1 and 5.9 summarises the key results and how they address the research question (RQ)s formulated in Chapter 1.

5.1. Results for RQ1 and RQ2

The results were analysed against the RQs and the corresponding hypothesis described in the previous chapter. Findings are summarised in Table 5.1.

Table 5.1: Results summary for RQ1 and RQ2

RQ	Hypothesis	Findings
RQ1	H_{1a}: Score AA > Score TA	Student achievement was significantly higher in authentic assessment for the composite group and groups isolated on independent variables.
RQ1	H_{1b}: Score AA_1 > Score TA_1	Student achievement was significantly higher in authentic assessment for the composite group and groups isolated on independent variables (except for student groups with more than three years of work experience).
RQ2	H_{2a}: Score AA_2 > Score TA_2	Student achievement was significantly higher in authentic assessment for the composite group and groups isolated on independent variables.
RQ2	H_{2b}: Score AA_2 > Score AA_1	Student achievement was significantly higher in the second authentic task for the composite group and groups isolated on independent variables.
RQ2	H_{2c}: Score TA_2 ~ Score TA_1	No significant difference in seafarer student achievement found for the composite group and groups isolated on independent variables.

The results are presented in the following section.

RQ1: Is there a significant improvement in seafarer students' academic achievement in authentic assessment when the scores are compared with traditional assessment scores?

Table 5.2 provides a result summary for RQ1.

Table 5.2: Result summary for RQ1.

Hypothesis	**Mean Score**	**Sample**	**S.D.**	**Effect Size**	**t-Test (df); P (two-tail) assuming unequal variance**
1a) Score AA > TA	AA (69.8%) > TA (52.5%)	AA (93) TA (96)	AA (14.6) TA (20.6)	0.98	t (172) 6.7; P <.05
1b) Score AA_1 > TA_1	AA_1 (63.8%) > TA_1 (52.4%)	AA_1 (93) TA_1 (96)	AA_1 (9.6) TA_1 (11.2)	0.55	t (184) 3.8; P <.05

Table 5.2 showed that AA significantly improved by 17.3% compared with TA; and AA_1 was 11.4% higher than TA_1. The hypotheses (H_{1a} and H_{1b}) designed for RQ1, thus, held true. In both hypotheses, the S.D. values indicated higher scattering amongst traditional assessment scores; and the effect size and the t-test values showed that the difference and variation in the scores were significant for reporting.

RQ2: Is there a significant improvement in seafarer students' academic achievement in formative authentic assessment when the scores are compared to summative traditional assessment scores?

Table 5.3 provides a result summary for RQ2.

Table 5.3: Result summary for RQ2.

Hypothesis	**Mean Score**	**Sample**	**S.D.**	**Effect Size**	**t-Test (df); P (two-tail) assuming unequal variance except ***
2a) Score AA_2 >TA_2	AA_2 (75.8%) > TA_2 (52.6%)	AA_2 (93) TA_2 (96)	AA_2 (7.3) TA_2 (10.7)	1.2	t (168) 8.7; P <.05
2b) Score AA_2 >AA_1	AA_2 (75.8%) > AA_1 (63.8%)	AA_2 (93) AA_1 (93)	AA_2 (7.2) AA_1 (9.6)	0.71	t (171) 4.8; P <.05
2c) Score TA_2 ~ TA_1	TA_2 (52.4%) ~ TA_1 (52.6%)	TA_2 (96) TA_1 (96)	TA_2 (11.2) TA_1 (10.7)	0.01	t (190) .11; P > .05 *assuming equal variance

Analysis of the composite scores as presented in Table 5.3 showed the following:

- AA_2 significantly improved by 23.2% when compared with TA_2;
- AA_2 significantly improved by 12% when compared with AA_1; and
- no significant difference was found between TA_1 and TA_2.

The hypotheses (H_{2a}, H_{2b}, and H_{2c}) designed for RQ2, thus, held true. In hypothesis H_{2a} and H_{2b}, the S.D. values indicated that the TA_2 and AA_1 scores were more widely scattered than the AA_2 scores; and the effect size and the t-test values showed that the difference and variation in scores was significant for reporting. In hypothesis H_{2c}, due to the

similarity in the composite mean score values of TA_1 and TA_2, as expected, the S.D. values indicated that the scores of both the traditional tasks were similarly scattered; and the effect size and the t-test values showed that the difference and variation in scores were not significant for reporting.

5.1.1 The effect of independent variables on students' academic achievement

H_{1a}: Score AA > Score TA

Table 5.4 summarises the effect of the independent variables on H_{1a}.

Table 5.4: Effect of independent variables on H_{1a}.

H_{1a}: Score AA > TA	**Mean Score AA > TA (%)**	**Sample AA/TA**	**S.D. AA/TA**	**Effect Size**	**t-Test (df); P (two-tail) assuming unequal variance**
Composite:	69.8 > 52.5	93/96	14.6/20.6	0.98	t (172) 6.8; P <.05
Work Experience:					
< 3 years	65.5 > 44.8	31/61	18.1/19.4	1.10	t (64) 5.0; P < .05
> 3 years	72.0 > 65.7	62/35	12.1/15.4	0.50	t (58) 2.1; P < .05
English (first language):					
Yes	79.3 > 64.0	33/27	9.8/15.3	1.20	t (43) 4.5; P < .05
No	64.7 > 48.0	60/69	14.3/20.7	0.95	t (121) 5.4; P < .05
Educational qualifications:					
High school	69.4 > 49.8	45/68	15.1/19.3	1.10	t (108) 6.0; P < .05
University	70.3 > 58.9	48/28	14.3/22.5	0.62	t (40) 2.4; P < .05

Table 5.4 showed that the scores, when isolated on the independent variables, revealed that the students' academic achievement significantly improved in AA when compared with TA. The hypothesis (H_{1a}), thus, held true for all the independent variables. The S.D. values indicated that the TA scores were more widely scattered than the AA scores; and the effect size and the t-test values showed that the difference and variation in scores was significant for reporting for all the independent variables.

H_{1b}: Score AA_1 > Score TA_1

Table 5.5 summarises the effect of the independent variables on H_{1b}.

Table 5.5: Effect of independent variables on H_{1b}.

H_{1b}: Score AA_1 > TA_1	**Mean Score AA_1 > TA_1 (%)**	**Sample AA_1/TA_1**	**S.D. AA_1/TA_1**	**Effect Size**	**t-Test (df); P (two-tail) assuming unequal variance**
Composite:	63.8 > 52.4	93/96	9.6/11.2	0.55	t (184) 3.8; P < .05
Work Experience:					
< 3 years	58.6 > 44	31/61	10.5/11.3	0.67	t (57) 3.0; P < .05
> 3 years	66.6 ~ 66.6	62/35	8.5/8.6	0.00	t (70) 0.0; P > .05
English (first language):					
Yes	73.6 > 64.2	33/27	7.9/9.4	0.54	t (51) 2.0; P < .05
No	58 > 47.6	60/69	9.4/11.1	0.51	t (127) 2.9; P < .05
Educational qualifications:					
High school	62 > 49.8	45/68	10.3/11.0	0.74	t (99) 3.0; P < .05
University	65.6 > 58.2	48/28	9.0/11.3	0.63	t (47) 1.5; P < .05

Table 5.5 showed that the scores, when isolated on the independent variables, revealed that the students' academic achievement significantly improved in AA_1 when compared with TA_1. The only exception was in the case of students with more than three years of work experience where the scores were found to be similar in value. This indicated that for the first task, traditionally assessed students with more than three years of work experience benefitted from their familiarity with the workplace, related the assessment task to the real-world context and hence, were able to respond as well as the authentically assessed students.

The hypothesis(H_{1b}), thus, held true for all the independent variables but with a single exception. The S.D values of TA_1 were more widely scattered than the AA_1 scores; and the effect size and the t-test values showed that the difference and variation in scores were significant for reporting in all groups isolated on the independent variables except for students with more than three years of work experience. Due to similarity in the AA_1 and TA_1 scores of students with more than three years of work experience, the S.D. values indicated similar scattering; and the effect size and the t-test values showed that the difference and variation in scores were insignificant for reporting.

H_{2a}: Score AA_2 > Score TA_2

Table 5.6 summarises the effect of the independent variables on H_{2a}.

Table 5.6: Effect of independent variables on H_{2a}.

H_{2a}: Score AA_2 > TA_2	**Mean Score AA_2 > TA_2 (%)**	**Sample AA_2/TA_2**	**S.D. AA_2/TA_2**	**Effect Size**	**t-Test (df); P (two-tail) assuming unequal variance**
Composite:	75.8 > 52.6	93/96	7.3/10.7	1.20	t (168) 8.7; P < .05
Work Experience:					
< 3 years	72.6 > 45.6	31/61	8.8/10.0	1.40	t (68) 6.6; P < .05
> 3 years	77.6 > 64.8	62/35	6.2/8.9	0.84	t (53) 3.7; P < .05
English (first language):					
Yes	82.8 > 63.8	33/27	5.2/8.8	1.30	t (41) 4.9; P < .05
No	71.4 > 48.2	60/69	7.4/10.6	1.50	t (122) 7.2; P < .05
Educational qualifications:					
High school	76.8 > 49.8	45/68	6.4/9.8	1.70	t (111) 8.8; P < .05
University	75 > 59.6	48/28	8.0/12.0	0.76	t (41) 3.0; P < .05

The scores, when isolated on the independent variables, revealed that the students' academic achievement significantly improved in AA_2 when compared with TA_2. The hypothesis(H_{2a}), thus, held true for all the independent variables. The S.D. values indicated that the TA_2 scores were more widely scattered than the AA_2 scores; and the effect size and the t-test values showed that the difference and variation in scores was significant for reporting for all the independent variables.

H_{2b}: Score AA_2 > Score AA_1

Table 5.7 summarises the effect of the independent variables on H_{2b}.

Table 5.7: Research findings to H_{2b}.

H_1: Score AA_2 > AA_1	**Mean Score AA_2 > AA_1 (%)**	**Sample AA**	**S.D. AA_2/AA_1**	**Effect Size**	**t-Test (df); P (two-tail) assuming unequal variance**
Composite:	75.8 > 63.8	93	7.2/9.6	0.71	t (171) 4.8; P < .05
Work Experience:					
< 3 years	72.6 > 58.6	31	8.8/11.3	0.69	t (57) 2.7; P < .05
> 3 years	77.6 > 66.6	62	6.2/8.5	0.74	t (113) 4.1; P < .05
English (first language):					
Yes	82.8 > 73.6	33	5.2/7.9	0.70	t (56) 2.8; P < .05
No	71.4 > 58	60	7.4/9.4	0.79	t (112) 4.3; P < .05
Educational qualifications:					
High school	76.8 > 62	45	6.4/10.3	0.88	t (74) 4.1; P < .05
University	75 > 65.6	48	8.0/9.0	0.55	t (93) 2.7; P < .05

Table 5.7 showed that the scores, when isolated on independent variables, revealed that the students' academic achievement significantly improved in AA_2 when compared with AA_1. The hypothesis (H_{2b}), thus, held true for all the independent variables. The S.D. values indicated that the AA_1 scores were more widely scattered than the AA_2 scores; and the effect size and the t-test values showed that the difference and variation in scores was significant for reporting.

H_{2c}: Score TA_2 ~ Score TA_1

Table 5.8 summarises the effect of the independent variables on H_{2c}.

Table 5.8: Research findings to H_{2c}.

H_1: Score TA_2 ~ TA_1	**Mean Score TA_2 ~ TA_1**	**Sample TA**	**S.D. TA_2/TA_1**	**Effect Size**	**t-Test (df); P (two-tail) assuming equal variance**
Composite:	52.6 ~ 52.4	96	10.7/11.2	.01	t (190) .11; P > .05
Work Experience:					
< 3 years	45.6 ~ 44	61	10.0/10.5	.08	t (120) .42; P > .05
> 3 years	64.8 ~ 66.6	35	8.9/8.6	.10	t (68) .42; P > .05
English (first language):					
Yes	63.8 ~ 64.2	27	8.8/9.4	.03	t (52) .09; P > .05
No	48.2 ~ 47.6	69	10.6/11.0	.02	t (136) .18; P > .05
Educational qualifications:					
High school	49.8 ~ 49.8	68	9.8/11.0	.00	t (134) .01; P > .05
University	59.6 ~ 58.2	28	12.0/11.3	.06	t (54) .22; P > .05

Table 5.8 showed that the scores, when isolated on the independent variables, revealed that there was no significant difference in students' academic achievement when TA_1 was compared with TA_2. The hypothesis (H_I: Score TA_2 ~ Score TA_1), thus, held true for all the independent variables. Although, the TA_2 score values was not always exactly equal to the TA_1 score values, the maximum difference between the two scores did not exceed 2% which was not considered significant in this research project. Due to the similarity in the mean score values of TA_1 and TA_2, as expected, the S.D. values indicated that the scores of both the traditional tasks were similarly scattered; and the recommended (Coe 2002) effect size and the t-test values showed that the difference and variation in scores was insignificant for reporting.

5.2. Results for RQ3

The results of the data analysis related to RQ3 is summarised for each stage of investigation in Table 5.9.

Table 5.9: Summary of results for RQ3.

RQ	Stage	Results summary
RQ1	Stage 1	Significant correlation found between seafarer students' perception of authenticity for the conceptually developed factor of transparency in criteria and their scores in the authentic assessment. Transparency of criteria was also found to be a significant predictor of student scores in the authentic assessment.
RQ1	Stage 2	Significant correlation found between seafarer students' perception of authenticity for Factors 2 (construction of knowledge) and 5 (transparency in criteria) extracted through a factor analysis and their scores in the authentic assessment. Factor 5 was also found to be significant predictor of student scores in the authentic assessment.

Reporting of the results in the following section is organized by each stage of data analysis.

5.2.1 Stage 1

The R-values for the correlation between the students' perception of authenticity (for the conceptually developed factors) in authentic assessment and their scores in the associated assessment task are detailed in Table 5.10.

Table 5.10: R-values of student perceptions of authenticity (conceptually developed factors) in authentic assessment and their scores in the associated assessment task.

	Relevance	***Task***	***Criteria***	***Trans Cri***	***Context***	***Constn Kn***	***Mul Opp***	***Totl Score***
Relevance	1							
Task	0.71	1						
Criteria	0.74	0.55	1					
Trans Cri	0.52	0.35	0.57	1				
Context	0.56	0.59	0.43	0.25	1			
Constn Kn	0.67	0.48	0.45	0.49	0.41	1		
Mul Opp	0.48	0.30	0.36	0.39	0.25	0.44	1	
Totl Score	0.17	0.19	0.13	**0.31**	0.10	0.18	-0.02	1

The R-values in Table 5.10 showed significant correlation (R-value higher than 0.25 was outlined in bold) between students' perception of authenticity for the factor transparency of criteria and their scores in the authentic assessment. Using the significantly correlated factor (transparency of criteria) and the scores in the authentic assessment, a linear regression analysis based on the recommended (Sarkar, Keskin & Unver 2011) confidence level of 95% (or p-value 0.05 or less) was conducted. Although confidence levels can be represented as 90%, 95%, 99% or any percentage (between 0 and 100%), the authors of this paper chose the most commonly used confidence level of 95% (Tan and Tan 2010). The findings of the regression analysis are detailed in Figure 5.1.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.3122596
R Square	0.097506
Adjusted R Square	0.0875885
Standard Error	13.967688
Observations	93

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1918.128195	1918.128195	9.831698956	0.002310037
Residual	91	17753.76428	195.0963108		
Total	92	19671.89247			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	51.299776	6.090641979	8.422720581	5.07168E-13	39.20146347	63.39808764	39.20146	63.39809
Trans Cri	5.2648653	1.679086156	3.135554011	**0.002310037**	1.929566951	8.600163712	1.929567	8.600164

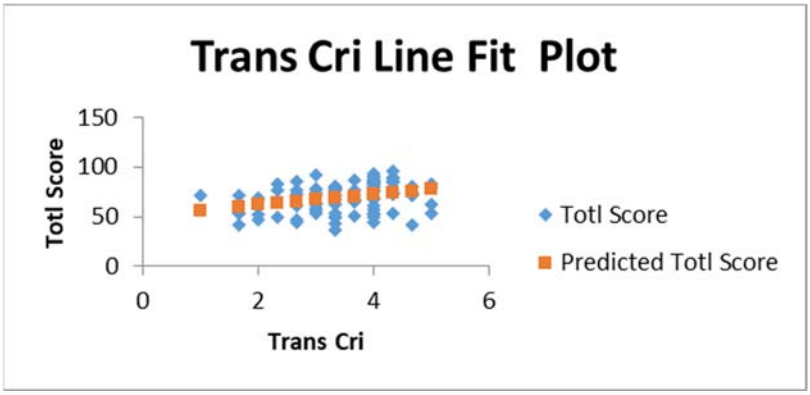


Figure 5.1: Regression analysis of seafarer students' perception of authenticity in transparency of criteria and their scores in authentic assessment.

The bold p-value (less than 0.05) of the factor transparency of criteria, as shown in Figure 1, revealed the factor to be a significant predictor of student scores in authentic assessment. However, this finding was based on a relatively lower value (8.8%) of the adjusted R-square.

5.2.2 Stage 2

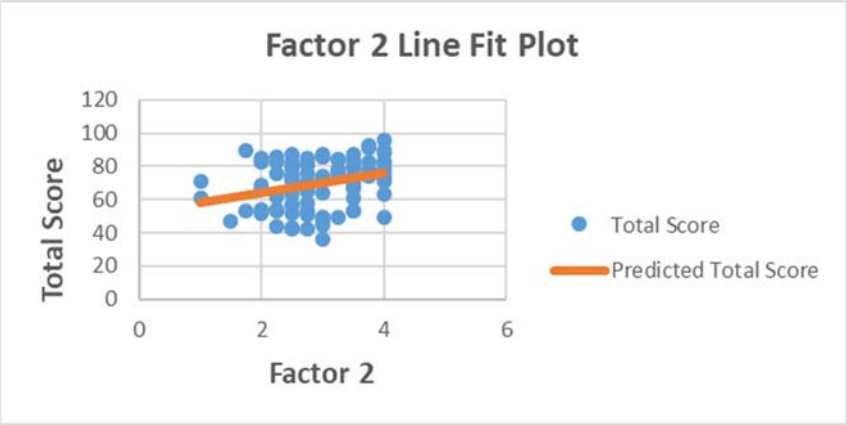
The R-values for the correlation between the students' perception of authenticity for the factors of assessment (extracted through factor analysis) and their scores in authentic assessment are detailed in Table 5.11.

Table 5.11: R-values of student perceptions of authenticity (factors extracted through factor analysis) in authentic assessment and their scores in the associated assessment task.

	Factor 1	***Factor 2***	***Factor 3***	***Factor 4***	***Factor 5***	***Totl Score***
Factor 1	1					
Factor 2	0.51	1				
Factor 3	0.77	0.60	1			
Factor 4	0.46	0.34	0.47	1		
Factor 5	0.41	0.50	0.56	0.19	1	
Totl Score	0.16	**0.29**	0.16	0.03	**0.31**	1

The R-values in Table 5.11 showed significant correlation (R-value higher than 0.25 was outlined in bold) between students' perception of authenticity for Factors 2 and 5 and their scores in the authentic assessment. Using the significantly correlated factors (2 and 5) and the scores in the authentic assessment, a multiple regression analysis based on the recommended (Sarkar et al. 2011) confidence level of 95% (or p-value 0.05 or less) was conducted. The findings of the regression analysis are detailed in Figure 5.2.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.351611748							
R Square	0.123630821							
Adjusted R Square	0.104155951							
Standard Error	13.84029554							
Observations	93							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	2	2432.052221	1216.026	6.348223	0.002635842			
Residual	90	17239.84025	191.5538					
Total	92	19671.89247						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	45.50989075	6.994083563	6.506913	4.23E-09	31.61492299	59.4048585	31.61492299	59.4048585
Factor 2	3.868048422	2.361499213	1.637963	0.104921	-0.823481943	8.559578786	-0.823481943	8.559578786
Factor 5	3.676608053	1.925711831	1.90922	**0.05942**	-0.149154615	7.502370721	-0.149154615	7.502370721



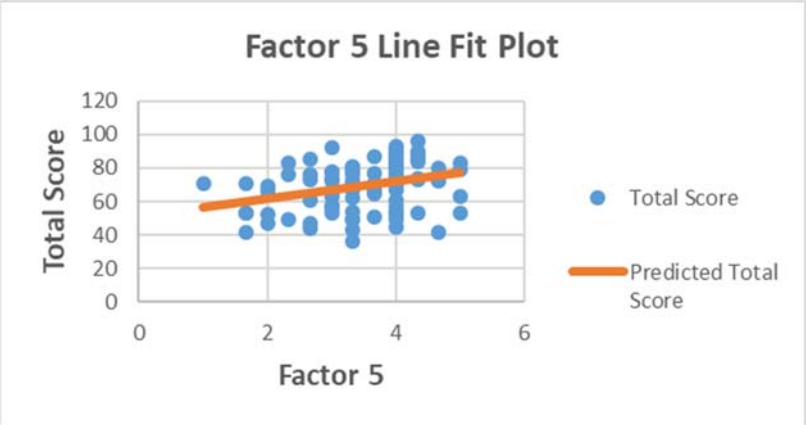


Figure 5.2: Regression analysis of seafarer students' perception regarding authenticity in Factors 2 and 5 and their scores in authentic assessment.

The bold p-value (less than 0.05) of the factor transparency of criteria, as shown in Figure 5.2, revealed the factor to be a significant predictor of student scores in authentic assessment. However, this finding was based on a relatively lower value (10.4%) of the adjusted R-square.

This chapter explained how the results of this study contribute to addressing each of the research questions formulated in Chapter 1. The previous results will be considered in greater detail in the next chapter of this thesis (Chapter 6).

6. DISCUSSION AND CONCLUSION

This chapter reflects on the findings (from the results discussed in Chapter 5) and their implications for research from a theoretical perspective, and also for practice. This chapter also draws attention to the novel contributions of this research study, the research limitations and constraints, and where areas of future research can develop and deepen the understanding of authentic assessment.

Section 6.1 explains how the key RQs formulated in Chapter 1 and reproduced in the following paragraph were addressed:

RQ1: Is there a significant improvement in seafarer students' academic achievement in authentic assessment when the scores are compared with traditional assessment scores?

RQ2: Is there a significant improvement in seafarer students' academic achievement in the formative authentic assessment when the scores are compared with summative traditional assessment scores?

RQ3: What is the correlation between seafarer students' perception of authenticity in assessment and their academic achievement in the associated assessment tasks?

Section 6.1 also discusses the implications of the research findings. Section 6.2 highlights the novel contributions of this research in theoretical, methodological, and empirical terms. Section 6.3 outlines the research limitations and constraints. Section 6.4 articulates the path to future research with concluding remarks in Section 6.5.

6.1. Research findings and their implications

6.1.1. Authentic assessment results in higher academic achievement for students

The results of this research project confirmed that seafarer students' academic achievement in authentic assessment is higher when compared with traditional assessment, and in doing so, addresses RQ1. This finding indicated that the academic achievement of students improved significantly when their responses to the questions in the assessment task were not relying on memorisation of information and imagining situations (as required in traditional assessment) but on the assimilation, integration, and analysis of information provided in a real-world context. The real-world context in which the authentic assessment tasks were based in this project mirrored tasks that were faced by seafarers during their professional careers requiring practical and realistic solutions.

The response to the questions in the tasks, then, reflected work or demonstrated knowledge and skills as would be required at the workplace. This enabled seafarer students to relate classroom learning to the work onboard ships, leading to higher academic achievement.

This finding, where seafarer students assessed authentically demonstrated higher achievement in comparison to traditionally assessed students, corroborated the findings of non-seafarer research (Brawley 2009; Schneider et al. 2002; Thomas 2000; Leon & Elias 1998; Gallagher, Stepien & Rosenthal 1992). Although, the findings were similar, this research made a unique contribution by studying participants in post-school settings compared with past research that was conducted in the educational settings of a school. For example, Brawley (2009) compared academic achievement of early childhood students, Schneider et al. (2001) for tenth and eleventh grade students, Thomas (2000) for tenth grade students, Leon and Elias (1998) for sixth grade students and Gallagher et al. (1992) for unspecified school students. Although authentic assessment has been implemented for student assessment in higher education (e.g. Gulikers 2006, Jonsson 2008, James and Casidy 2016), the literature review conducted for this research project did not find any studies comparing students' academic achievement between traditional and authentic assessment in post-school settings.

This research also distinguishes itself from past research by using two separate student groups as the 'control' and 'treatment' group. Past research (e.g. Brawley 2009) used the same student group for both traditional and authentic assessments. In cases where the same group of students are used for both assessments, the higher achievement of the students transitioning from traditional to authentic assessment may be partially attributed to the 'learning effect', i.e. the gain in student knowledge that may have occurred in the time between the administrations of the two assessments. Learning effect creates an additional variable, which was avoided in the approach designed for this project.

Although, past researchers such as Schneider et al. (2002), Thomas (2000), Leon & Elias (1998), and Gallagher et al. (1992) used two independent randomly assigned groups for comparison between authentic and traditional assessment performance, additional variables other than the 'authentic' design of the assessment may have been introduced due to the nature of the tasks or associated learning. For example, Leon & Elias (1998) used 'portfolios' versus 'self-selected performance-based projects; Gallagher et al. (1992) used 'open-ended questions' versus 'authentic performance task'; and Schneider et al. (2002) and Thomas (2000) used two separate groups with a different learning experience before the authentic and traditional tasks were administered. In contrast, the focus of this research was only on studying the impact of the 'authentic' design and the 'nature of implementation' (as the differing aspects between the two types of

assessment) on students' academic achievement keeping the remainder of the variables (e.g. learning content, assessment questions and duration, etc.) constant.

6.1.2. Formative assessment results in higher academic achievement for students

Hattie (2019; 2009) ascertained the major influences on student achievement by synthesising more than 800 meta-analyses in education. He reported that one of the key requirements for skills improvement is feedback on the students' current level of skills and multiple opportunities to practice the skills, therefore making the learning pathway 'visible' to the learner. The findings of this thesis concurred with Hattie and in doing so, also addresses RQ2. The formative authentic assessment employed in this research project provided students with an opportunity to receive individual feedback on their performance in the first authentic assessment task (AA_1) before attempting the second authentic assessment task (AA_2). Feedback on a students' current ability to perform an assessment task and providing suggestions to improve and attain expected levels, encourages students to take necessary actions to close the gap in their ability (Zhang & Zheng 2018). Higher academic achievement in AA_2 as compared with AA_1 indicated that one of the reasons for the improvement in achievement may be attributed to the feedback that allowed the seafarer students to recognize the gaps in their knowledge, re-evaluate their learning approaches and implement new strategies to improve their scores. In comparison to the formative assessment, the feedback obtained by the students in the summative traditional assessment task proved to be too late for the control group of seafarer students to make any adjustments to their learning process to improve their scores.

Findings of this research project thus indicate that in the context of seafarer education, a shift is required from summative and traditional oral assessments that declare students as 'fail' before being provided with a feedback or another opportunity. The use of summative examinations at the end of the learning period represents the final judgement on the students' performance and is often too late to make any changes to the learning strategies. In this mode of assessment, borderline students are offered 'supplementary' examinations but without necessarily understanding their learning gaps. The alternative approach advocated in this thesis would see educators provide timely and efficient feedback to students; and receive counter feedback to reflect on areas they may improve upon as well. Students should take advantage of the feedback and work closely with the educators and assessors to become active participants in the learning process by recognizing their strength and weaknesses; and in establishing realistic learning goals. This develops their metacognitive ability of reflecting on their current learning practices and improving on them. Reflection on practices is a critical part

of professional performance required to avoid errors and develop the essential competencies to perform the workplace tasks.

6.1.3. Factors that influenced the academic achievement of students

6.1.3.1 The influence of work experience on students' academic achievement

In this research project, the analysis of the student scores within the control and treatment groups (overall and when isolated on independent variables) revealed that academic achievement was significantly higher for students with more than three years of work experience as compared with students with less than three years of work experience. This implied that students with lesser work experience may not have had adequate exposure at their workplace in performing tasks similar to the assessment tasks designed for this project which also affected their ability to perceive the authenticity of the assessment. In comparison with the lesser-experienced students, there is a possibility that students with more than three years of work experience were more familiar with the assessment tasks and had performed them in the workplace contexts, which enabled them to significantly score higher than their less experienced counterparts.

Also, greater work experience provides the students with a greater ability to think, analyse, and develop solutions not only for problems they are familiar with, but non-familiar issues related to the workplace as well (Chaudhry & Rasool 2012). This was evident in the scores of seafarer students in the comparison of AA_1 to the first traditional assessment task (TA_1) and AA_2 to the second traditional assessment task (TA_2), when isolated for the independent variable of 'more than three years of work experience'. The scores of AA_1 were similar to TA_1 when isolated for the specified variable which indicated that seafarer students with higher work experience may negate the advantage provided through the real-world context of authentic assessments due to their experience in performing similar tasks in the workplace.

However, the scores of AA_2 were significantly higher than TA_2 when isolated for the same variable. This suggested that the factor of 'higher work experience' could not nullify the advantage provided through the real-world context in the second authentic assessment task. Since, the comparison was between the same group of students for both tasks, the only advantage provided to the authentically assessed students over the traditionally assessed students for the second assessment task was the authenticity of the assessment, a feedback and an opportunity to improve on their performance and resulting scores.

Educators face the challenge of teaching and assessing students with differing work experience within the same cohort. Although seafarer students may have differing work experiences when they enrol for their studies, on completion and receiving the CoC they are all expected to perform to the minimum and global professional standards (and beyond) as required for their certification and described in the STCW Code. Ensuring students can perform the workplace-based authentic task at least to the minimum standards (as described in the assessment rubric) before they are deemed competent will ensure minimum acceptable or required standards have been achieved before awarding the CoC. Moreover, formative implementation of authentic assessment will further ensure reliability in performance and multiple evidence of competence. Adhering to minimum competency standards will ensure graduating students' have acquired the minimum required competence irrespective of the differing work experience. This would be a distinct advantage to the IMO who introduced the STCW Code to achieve global consistency in competency standards. However, educators should be mindful about the difference in student ability due to differing work experience and strive for parity via greater opportunities to practice similar tasks before the main assessment (Lenz et al. 2000). Teacher feedback on practice attempts will allow students to identify their areas of weakness and address them. Due to time constraints, one of the limitations of this project was its' inability to provide students with an opportunity to practice similar tasks before the main assessment.

6.1.3.2 The influence of proficiency in the English language on students' academic achievement

Analysis of the student scores also revealed that student achievement was significantly higher for students with English as their first language as compared with their non-native English-speaking counterparts within both the control and treatment groups (overall and when isolated on independent variables). One of the key reasons for this finding may be attributed to the format of the assessment which required students to respond to questions based on a case-study. Answering the questions in English may have affected the performance of the students who were not proficient in the language, and hence, lowered their academic achievement.

In countries where training and assessments are conducted using the English language (for example, in Australia where this project was conducted), a key implication for educators lie in facing the challenge of teaching and assessing students with differing abilities in communicating using the English language. This was also pointed out in the article by Kainth (2019) that discussed the Victorian government's push for a review of English language requirements for international students amid concerns they are falling behind due to a lack of English skills. One of the ways educators may seek to achieve

parity is by laying out minimum requirements, for example through the International English Language Testing System (IELTS) score. This is already the case with the seafaring course at AMC which requires international students (especially those originating from non-English speaking countries) to demonstrate the ability to achieve a minimum IELTS score. Also, with English being the lingua franca of the sea, seafarer educators should examine the possibility of raising standards in this area and developing educational programs to assist students to meet the raised standards.

However, based on the current scenario of a classroom with diverse population of students with different ability to communicate in the English language, educators could also investigate ways to design authentic assessments that require students to perform tasks that require more hands-on approach with lesser focus on language abilities. Another solution could also be to involve students in the design of the authentic assessment. This may be achieved by familiarising the students with the assessment rubric and seeking their inputs on how an assessment task may be designed to not only be authentic to their workplace settings but also allow the testing of competencies listed in the rubric. Including student voice will address respective concerns and allow educators to plan for them in advance.

6.1.3.3 The influence of educational qualifications on students' academic achievement

Analysis of the scores (overall and when isolated on independent variables) within the treatment group revealed that the educational qualifications of a student had no significant impact on student achievement in authentic assessment. For example, in both the first and second authentic assessment tasks, students with university level of education did not score significantly higher (a comparison of means indicated that the difference of marks was less than two percent) than the students with only high school qualifications. This finding indicated that the authenticity of assessment evened out student scores that reflected their ability to analyse assessment scenarios for critical assimilation of information towards providing response to assessment questions. For example, the student responses in authentic assessment were not only based on their ability to read and comprehend a case-study but also on the cues provided through the immersive and authentic, real-world demonstration of the case study that engaged all the sensory perceptions of the students. According to the Atkinson-Shiffrin model for memory, vivid cues provided through experiences that trigger the sensory registers assists in the retrieval of information and prevent lapses in memory (Atkinson & Shiffrin 1968). Hence, all authentically assessed students were able to retrieve the information provided to them through the real-world demonstrations and answer the questions asked in the case study.

In comparison to authentic assessment scores, analysis of the scores within the control group (overall and when isolated on independent variables), revealed that students with university level of qualification scored significantly higher than students with only high school qualifications. In absence of cues provided through the immersive demonstrations of the real-world contexts, students in traditional assessment relied on their ability to analyse information based on their ability to read and comprehend a descriptive case-study. Hence, students with university qualifications used their academic experience of participating in similar context-devoid assessments and scored over their lesser qualified (high school) counterparts.

However, it must be acknowledged that the investigators of this research study did not enquire on the different kinds (country of issue, university of study, etc.) of under-graduate and post-graduate qualifications the research participants claimed to possess. Since, it was not possible to determine the quality of university education the research participants may have experienced prior to this research study, the findings may be contextualised to this research only. Although, it is likely that the relationship between educational attainment and academic scores in assessment is less than perfect, educators must investigate ways to design authentic assessment to bridge the important gap between students with differing educational backgrounds.

6.1.4. Transparency of assessment criteria is a significant predictor of academic achievement

In their research, Gulikers (2006) and Jonsson (2008) reported that the timing and transparency of assessment criteria enhanced student achievement in authentic assessment amongst their research participants. The results of this thesis study showed that the student perceptions of authenticity for the factor transparency of criteria correlated most significantly with their academic achievement, and in doing so, concurred with Gulikers (2006) and Jonsson (2008); and addressed RQ3.

Transparency of assessment criteria is essential for learning (Biggs & Tang 2011; Reddy 2007; Wiggins 1989) and providing the criteria at the beginning of the learning period (thus making the assessment transparent) is an essential and key requirement for authentic assessment (Villaroel et al. 2018). The finding of this thesis study indicated that seafarer students had significantly higher achievement when they found the assessment criteria to be transparent. Having the assessment criteria (detailing standards of performance) beforehand provided a roadmap of the content to be learned, while allowing the students to construct the understanding of the topic. This project provided the assessment criteria at the beginning of the learning period through the use of assessment rubrics.

Although, the rubric was provided to the control as well as the treatment group, the real-world scenarios demonstrated in the authentic assessment allowed the authentically assessed students to analyse the scenarios and construct responses towards the achievement of the standards described in the rubric. For example, when the students were asked to 'recognize all the barriers to effective communication' (as specified in the assessment rubric) in the scenario described in the case study, the authentically assessed students were able to experience the wind and rain that hampered communication in the emergency scenario. In comparison to the authentically assessed students, the traditionally assessed students were unable to recognize the same barriers to communication from the descriptive case studies absent of demonstrated scenarios. The authentically assessed students, thus, used the rubrics to reflect on their learning and carry out self-assessments of their thinking and practices towards achievement of the required standards.

6.1.5 Enhancing robustness of research findings through a larger sample size

Past research findings that provided empirical evidence on the impact of authentic pedagogical practices on outcomes outlined in this paper (student engagement, ability to transfer skills, etc.) using qualitative methodology have been run with relatively low sample sizes. For example, past research investigating the impact of authentic pedagogy on student engagement was based on 6 students (Richards Perry 2011), 11 students (Quartuch 2011), 4 students (Findlay 2013), and 10 students (Morrissey 2014). Although Morrissey (2014) justified the use of qualitative methodology (interviews) on a low sample as a means to gain an in-depth understanding of the concerned research phenomena, these methods do not replace sound quantitative research. The research presented in this thesis contributed through the robustness of the findings in the area of authentic assessment based on a relatively larger sample of 93 students. The author argues that this provides a strong level of reliability and generalisability from the results, while still recognising the need for further research in this area.

6.1.6 Low value of adjusted R-square

The findings of the regression analysis presented in this paper, are based on a relatively low value (8.8% in stage 1 and 10.4% in stage 2) of adjusted R-square. The adjusted R-square value focuses on explaining the observed variation in the dependent variable due to the independent variable (Lukacs, Burnham & Anderson 2010). This implied, that the significant factor (transparency of criteria) in this paper, although important, did not explain the majority of the variance in the student scores. This was also evidenced by the fact that Factor 1 accounted for the majority of the variance (38.5%) and, did not

correlate significantly to the scores. Hence, it was a possibility that the correlation and regression model adopted in this paper may not have included important factors of assessment before measuring the independent variable of perception of authenticity in assessment. For example, factors of assessment such as collaborative assessment (Ashford-Rowe, Herrington & Brown 2014; Gulikers 2006), student ownership of task design (Gulikers 2006), completion of task and collating of evidence competence by students over a sustained period (Morrissey 2014); and presentation of student work to an audience (Herrington 1997) were rejected at a theoretical level due to the following reasons:

- Collaborative authentic assessment was rejected since the research by Gulikers (2006) revealed that students and teachers rated this factor (described as 'social context') as the least important dimension of authentic assessment. Moreover, demonstrating individual competence in the units of learning is essential for seafarer certification (International Maritime Organization (IMO), 2011).

- Factors such as collaborative authentic assessment, student ownership of task design through portfolio-based authentic assessment, and completion of task over a sustained period of time via student-selected authentic-based projects were also rejected to avoid plagiarism in student work. This research required seafarer students to complete the assessment task under the supervision of externally employed invigilators. The factors were also rejected since inclusion of these factors in the assessment design would have created uncontrolled additional variables (e.g. variation in student groups, variation in task design and variation in time taken to complete task) other than the authentic design that would have impacted student performance.

- The factor requiring presentation of student work to an audience was rejected since it was incongruous to the nature of the assessment task developed for this paper.

The relatively low value of adjusted R-square may have also resulted from the use of the quantitative survey to measure student perceptions. This is because the use of Likert scales in the quantitative survey may have limited the students from outlining, describing, and adequately conveying the other factors of authentic assessment that they perceive to be significant towards obtaining a higher academic achievement. Instead, the students were compelled to choose the significant factors amongst the choices provided through the survey which may have led to an inadvertent omission of factors. This was also evidenced by the perception study by Gulikers (2006) in which, the

quantitative data did not reveal an overall differing perception of authenticity in task, but the qualitative investigation revealed otherwise.

Goodwin & Leech (2006) recommended examining the variability in the data (dependent and independent variable) if the resulting correlation was lower than expected. Lack of variability (indicated through low values of standard deviation) lowers the correlation value between variables (Goodwin & Leech, 2006). To examine the variability, this research calculated the standard deviation values for the student survey responses for perception in authenticity (independent variable) and the composite student scores (dependent variable). The standard deviation for student scores was 14.6 (mean score 69.8/100; minimum 36/100; maximum 96/100). The standard deviation, thus, indicated a relatively low value, which may have contributed to the lower correlation between the variables. Similar to the dependent variable, the standard deviation values of the student responses to the perception survey (as shown in Appendix 7) had relatively low values, which may also have contributed to the lower correlation.

Lack of variation in student scores indicated evenness in student performance. This may imply that the evenness in performance may have resulted due to the transparency in assessment criteria that provided all students with the same guidelines to obtain higher academic achievement; and the formative authentic assessment that provided students with more than one opportunity to do so. This argument is based on past researchers (Black & William 1998a, 1998b; Sadler & Good 2006; Jonsson 2008), who claimed that in studies characterised by formative assessments and transparent criteria, the difference in student achievement between high- and low-performing students is typically reduced.

6.2. Research contributions

After having discussed the results from the data analysis in Section 6.1, this section compiles the discussion with the findings presented in all the previous chapters to highlight the novel and key contributions of this research and how they addressed the research gaps identified in this thesis. In doing so, this section underlines the theoretical, methodological, and empirical contributions to the area of authentic assessment and SET.

6.2.1. Theoretical Contribution

6.2.1.1 Reviewed the STCW Code through the lens of Outcomes-based education (OBE)

The IMO revised the STCW Code'78 through the 1995 (since known as STCW'95) and 2010 amendments intending to fundamentally improve the knowledge-based training mandate by making it outcome-based. As a requirement of OBE and for the purposes of the certification and licensing, seafarers are required to demonstrate the achievement of the STCW standards through assessments. The 'standards' in the STCW code were expected to act as guidelines for regulatory bodies and SET worldwide to develop consistent and uniform training outcomes. A critical review of specific excerpts from the STCW code showed that the code largely fails to provide a 'standard' that can assure assessment of seafarers to one of the most critical outcomes: the performance expected at a level of work in the industry.

Various examples from the code were used to show how it fails to provide explicit guidelines for outcome development, contexts for assessment, and standards and criteria for student performance. Based on the critical review and examples, this paper argued that different ideas as to 'outcomes' has been confusing the interpretation of STCW and, therefore, how seafarer students are being assessed. It showed that the STCW code is too vague and this may lead to individual interpretation in adopting learning and assessment processes towards competence development, which creates the risk of seafarers graduating with CoCs but lacking the required competence for workplace operations.

Hence, this research made a unique review of the STCW code through the lens of the OBE to investigate if the Code fulfils the requirements of the latter. This research, thus, provided a review-based evidence to the IMO, as the developer of the STCW code, to rethink and improve on the document that guides maritime regulators and educators/assessors in SET for a seafarer's competence assessment.

6.2.1.2 Redefined authentic assessment

This research did not simply borrow the definition of authentic assessment from past papers, instead it was redefined by collating the characteristics of the assessment provided by most commonly cited authors in the area. This research provided a unique definition of authentic assessment which herein will encompass the following:

> **tasks** resulting in outcomes in a real-world context that require the integration of competence to solve forward looking questions and ill-structured problems;

processes that require performance criteria to be provided beforehand and evidence of competence to be collected by the student; and
outcomes that result in valid and reliable student performance, contextual and multiple examples of evidence of competence, higher student engagement, and transfer of skills to different contexts.

Such a definition, that was inclusive of all the essential characteristics of the assessment, was previously missing from the area of authentic assessment. Redefining authentic assessment in terms of tasks, processes, and outcomes provides a more holistic understanding of the concept for theory and practical implementation for educators and assessors. The new definition does not simply focus on the inclusion of a real-world context to establish an assessment as an authentic assessment but also on other characteristics (forward-looking questions, ill-structured problems, etc.), which should be included in the design of authentic assessment. Understanding the new definition will dismiss the popular notion of maritime educators and assessors who believe that the use of simulators results in the authentic assessment of seafarer students. Simulator tasks provide a real-world context for skills and knowledge application but if they not designed with ill-structured problems that require an integration of competence, conditions of authenticity are not met. Similar to simulators, understanding the redefined concept of authentic assessment will also promote improved assessment design for practical (e.g. fire-fighting) and laboratory (e.g. machinery) exercises.

6.2.1.3 Revealed gaps in past approaches to validity and reliability of authentic assessment

Through extensive literature reviews described in Chapter 2 of this thesis, this research revealed gaps in past approaches to addressing the validity and reliability of authentic assessment when implemented with and without the use of the assessment rubrics. The two separate literature reviews (n=236 articles) conducted for this thesis found that past implementation of authentic assessment had typically addressed none or few aspects of validity and reliability only. Reliability and validity are crucial to the acceptance of authentic assessment as an accurate measure of knowledge, skills, and behaviours (Stevens 2013). There are numerous extraneous variables that affect the validity and reliability of the rubrics when used an assessment instrument (Taylor 2011).

If these variables are not addressed, then the validity and reliability of the assessment and the resulting outcomes becomes questionable (Olfos & Zulantay 2007). For example, a lack of construct validity (as found in most cases) may indicate that that underlying psychological variables such as problem-solving, social interaction, and communication,

which are required universally in most professions, were not adequately assessed in these cases.

6.2.1.4 Established a hypothesized relationship between the validity and reliability of authentic assessment and its resulting outcomes in seafarer education

A review of the current traditional assessment methods used to assess a seafarer's competence was conducted as part of this research study. To bridge the resulting gap in the students' expectations with the assessment process (as highlighted in the review), the need for improvement in the process was established in this research study. Authentic assessment was introduced as a possible alternative and in this regard, a theoretical definition of the assessment was formulated in this research study. According to the definition, the tasks and processes of authentic assessment should result in the outcomes of: higher student engagement, ability to transfer skills to different contexts, contextual and multiple examples of evidence of competence, and valid and reliable student performance. The resulting outcomes, which were also justified through the provision of empirical evidence from past research, would theoretically meet student expectations with the assessment process. However, to ensure that the 'authentic' tasks reflect workplace situations requiring students to apply knowledge, skills, and behaviours to professional standards; and to test consistency of such performances, authentic assessments and the resulting performances should be judged by the essential criteria of validity and reliability.

Building on existing research, this thesis made a theoretical contribution in the area of authentic assessment via a hypothesized relationship, that is 'if aspects of validity and reliability of authentic assessment are improved holistically, then assessment of seafarer education and the resulting evidence of student competence to perform workplace tasks can be significantly improved'. The significant improvement will be in comparison with an authentic assessment that has not addressed any or typically one or two selected aspects of validity and reliability. Thus, holistic improvement in the validity and reliability of authentic assessment will crucially, raise the positive perceptions of students and employers with regard to the resulting assessment outcomes, assuring that the assessment is to a standard they can 'trust'. Using theories of learning and empirical evidence from past research, this research connected different aspects of validity and reliability to authentic assessment outcomes. Educators looking to improve any of the mentioned outcomes at their institutes can use the relationship established in this research to make changes in their assessment practices.

6.2.2. Methodological Contribution

6.2.2.1 Developed a conceptual framework (AAFSET) towards addressing upon the different aspects of validity and reliability of authentic assessment

This research made a methodological contribution by developing a novel, conceptual and practical framework (AAFSET) that promoted a holistic approach to authentic assessment, thereby providing greater assurances of validity and reliability throughout all stages of assessment within seafarer programs. The framework addressed the various aspects of validity (content, construct, and criterion) and reliability (internal consistency, inter-rater, split-half, and test-retest) during the different stages (before, during, and after) of the implementation of authentic assessment. The AAFSET framework that incorporates a feedback loop will use valuable data from student performances, and student and employer perceptions to enhance validity and reliability of authentic assessment and its resulting outcomes. Maritime educators and assessors intending to implement valid and reliable authentic assessment, may use this framework as a guide during different stages of assessment implementation. The framework will provide guidance to ensure that the essential constructs of a students' competence is being assessed; and the resulting evidence of competence meets the standards expected at the workplace and the statutory requirements (STCW Code) of seafarer assessment.

6.2.3. Empirical Contribution

6.2.3.1 Contributed empirical evidence regarding authentic assessment and its impact on seafarer education

Extensive literature reviews conducted as part of this project revealed a global absence of empirical evidence on authentic assessment in seafarer education. This research investigated the difference in seafarer students' academic achievement comparing traditional and authentic assessment scores. Moreover, through the correlation analysis, the research presented in this thesis revealed the factors of assessment that had a significant relationship with seafarer students' academic achievement. The factors that correlated significantly not only explained higher academic achievement in authentic assessment, but also provided valuable insights into seafarer assessment design, where the inclusion of the highly correlated factors in assessments may lead to improved student scores in the future. Hence, this research made a valuable contribution through the collection of empirical evidence towards the impact of authentic pedagogical practices in seafarer education and addressed research gap 1.3.3 identified in this research study.

6.3. Limitations and constraints

6.3.1 Relatively low value of adjusted R-square

One may argue that a key limitation of this paper was that the findings of this project were based on a relatively low value of adjusted R- square. The adjusted R square value focuses on explaining the observed variation in the dependent variable due to the independent variable. However, the focus of this paper was not to explain variation, but to find an association through correlation between the independent variable (perception of authenticity) and dependent variable (scores). In this context, the adjusted R-square value was irrelevant; and the low R- square value with statistically significant parameters was more valuable than a high R-square value accompanied with statistically insignificant parameters.

6.3.2 Inability to provide students with opportunities to practice assessment tasks

This research highlighted that educators face the challenge of assessing students with different work experiences, proficiency in the English language, and educational qualifications. The authentic assessment employed in this project was able to achieve parity in performance and resulting scores only between students with university and high school qualifications. In all other cases of authentic assessment, and for the traditionally assessed students, academic achievement was higher for students with higher work experience, proficiency in the English language, and educational qualifications. To address the needs of the learners with different backgrounds and achieve equity in academic achievement, this project recommends educators to provide students with the opportunity to practice tasks similar to the assessment tasks. Inability to do so due to time constraints, is one of the key limitations of this research.

6.3.3 Research findings contextual to selected unit

The data for this research study was collected from the seafarer educational programme conducted at AMC due to the convenience in the selection and recruitment of participants. The data was collected for the selected unit of Managerial and leadership skills since this unit provided the maximum number of enrolments, and hence, possible participants. Although, the sample size was relatively larger than the past qualitative research investigations in the area of authentic assessment and an attempt was made to engage the maximum sample size, findings may have differed if a larger sample size was available. It must also be acknowledged that the selected unit pertained to a non-technical skill of seafaring and hence the assessment task was designed to assess the

desired competencies of managerial and leadership skills. Results may have also differed if a technical unit (e.g. celestial navigation and ship stability) that required an assessment of a different set of competencies (math skills, use of ship's equipment, etc.) was selected.

6.3.4 Collaboration required between cross-disciplinary teams

Authentic assessment implemented in this research required the assistance of additional staff members employed at AMC to demonstrate and create the real-world contexts. This suggested that embedding authentic assessment in a course may require cross-disciplinary teams to work closely together. This differs from the current work allocation methods of one lecturer per subject. Additionally, to educators working together, policies of education institutes need to support the organization of required resources towards assessment implementation. For example, this research project required resources such as lifeguard, swimming pool, liferafts, and smoke, rain and wind generators. The lack of or restricted resources may create challenges for assessors in reproducing contexts for assessment that resemble the real-world. For example, in this research project, the real-world context may have been more authentic if access to ships with the required facilities to demonstrate the case study were available. Although, AMC has access to a training vessel, large student numbers as research participants did not allow the assessment to be implemented on the vessel due to restrictions on number of persons allowed on ships for safety reasons.

6.3.5 Authentic assessment is resource-intensive

Although the implementation of authentic assessment in this project provided a significant advantage through higher academic achievement for seafarer students, the discussion section on this project would not be completed without the inclusion of the analysis conducted on the disadvantages of implementing authentic assessment. Past research (Neely & Tucker 2012; Wiggins 1989) suggested that the direct and indirect costs of developing authentic assessments is twice as much as traditional assessments. Day et al. (2018) claimed that authentic assessments are often time consuming. Hence, a comparative cost and time analysis for the resources used in developing the traditional and authentic assessment was conducted for this research. Table 6.1 details the differences in the resources used and the costs incurred.

Table 6.1: Comparison of resource estimation for the implementation of traditional and authentic assessment.

Item Number	Resources	Traditional Assessment Costs	Authentic Assessment Costs
1.	Staff	1 lecturer = \$180/hour 2 invigilators x \$40 = \$80/hour **Total cost (TA_1 and TA_2) = \$260**	1 lecturer = \$180/hour 4 staff members to demonstrate authentic case studies x \$180 = \$720 2 invigilators x \$40 = \$80/hour Cost for AA_1 = \$980 Cost for AA_2 = \$980 **Total cost (AA_1 and AA_2) = 2 x \$980 = \$1960**
2.	Classrooms	1 classroom (TA_1 and TA_2) x \$100/hour = \$100 **Total cost (TA_1 and TA_2) = \$100**	1 classroom (AA_1) x \$100 = \$100 1 classroom (AA_2) x \$100 = \$100 **Total cost (AA_1 and AA_2) = \$200**
3.	Facilities	**None**	AMC Pool; liferaft; smoke generators; safety equipment = \$600/day Lifeguard = \$40/hour Cost for AA_1 = \$640 Cost for AA_2 = \$640 **Total cost (AA_1 and AA_2) = 2 x \$640 = \$1280**
4.	Time	1 hour for implementing both tasks; Both tasks completed in one day; 2 hours for providing students with common feedback on their performance; 2 hours of assessor's meeting **Total time used = 5 hours**	2 hours for developing and demonstrating both authentic case studies; 2 hours for implementing both tasks; Both tasks completed in three weeks; 4 hours for providing students with common feedback on their performance; 4 hours of assessor's meeting **Total time used = 12 hours**
5.	Answer booklets	100 booklets (TA_1 and TA_2) x \$0.70 = \$70 **Total cost (TA_1 and TA_2) = \$70**	100 booklets (AA_1) x \$0.70 = \$70 100 booklets (AA_2) x \$0.70 = \$70 **Total cost (AA_1 and AA_2) = \$140**

In Table 6.1, the time and cost (shown in Australian dollars) analysis showed that in this project, the cost of implementing a new and innovative assessment (authentic assessment) was significantly more than maintaining an existing assessment (traditional assessment). For such cases, Joughin, Dawson & Boud (2017) argued that changes in the assessment regime become justified if the benefits outweigh the costs. This project provides the evidence of benefit through higher academic achievement for authentically assessed students, and in doing so justifies its implementation in SET.

6.4. Future research

6.4.1 In-depth analysis of significant predictors of academic achievement

While the quantitative method used in this research created significant strengths, there are also generalised weaknesses or limitations associated with using these tools to understand human constructs. The quantitative methodology used a survey based on a Likert scale that limited the response of the seafarer students to a perception survey. Certain variables (collaborative assessment and student ownership of task design) may have been rejected at a theoretical level and intentionally omitted from the data analysis model used in this project. Future research should investigate seafarer students' perceptions with regard to authenticity in assessments when compared to the actual workplace through the use of qualitative methodologies, such as interviews, focus groups etc. The use of qualitative methodologies may allow the researcher to probe further and the research participants to provide with, not only an explanation of their perceptions of authenticity with factors of assessment that significantly affect their academic achievement, but also ways to improve the authenticity of the factors.

To gain more insight into seafarer students' perception regarding authenticity of assessment, future research should also investigate variations in student perceptions of different kinds of authentic assessment. The impact of the different kinds of authentic assessment on student learning should then be investigated using the techniques of observation and monitoring to gain an understanding of the strategies adopted by the students to cope with the differing authentic assessments.

6.4.2 Longitudinal study investigating the impact of authentic assessment on seafarer students' competence

Authentic assessments conducted in real-world contexts should ideally prepare students to perform similar tasks at the workplace. This research study investigated the seafarer students' ability to perform the assessment task in the educational settings of AMC but was unable to do so in the workplace settings of a ship due to logistical constraints. Hence, future research should design a longitudinal study that tracks seafaring students into the workforce to investigate if they perform better after being assessed through authentic assessment compared with traditionally assessed students.

6.4.3 Replicating research study in other SET institutes and for other units of study

The data for this research projected was collected in the educational settings of AMC due to the easy accessibility to the research participants. However, it is possible that results of this study may differ if conducted in a different environment of training and the chosen unit of study. Hence, similar studies should be replicated in other SET institutes or VET settings globally, and for other units of study to compare findings.

6.5. Conclusion

The current research is grounded in the assertion that student assessment should not only be a formality conducted as an educational practice but integrated into a carefully planned and implemented process to serve the focal goal of improving student learning. In the context of seafaring education, assessment of seafarers forms the basis on which they can be granted a CoC for particular roles and levels of responsibility, providing them with a licence to operate ships and its' equipment. Ships are "floating structures" at sea which not only carry human resource, but in many cases, different volumes of cargo, oil and other entities. Past records indicate that accidents involving ships have often resulted in loss of lives and damage to the marine environment, affecting not only the stakeholders involved in the shipping industry but sometimes the entire community. Analysis of shipping accidents has proved that a major percentage of these accidents are caused by operational mistakes and errors by the ships' crew. Seafaring industry will always be at risk of major operational errors if competence levels of seafarer operators are not accurately and adequately assessed before issuing them with the CoC. The STCW Convention led to the development of the STCW Code to provide global, minimum standards of competence assessment for seafarers. The outcome of the assessments, in the form of a CoC, provides justification for its holder to seek job opportunities; for the employers to recruit, reward and train; and for the maritime regulators to form a workforce that comprises of professionals with standardized competence levels.

The 'standards' in the STCW Code were expected to act as guidelines for regulatory bodies and MET providers worldwide to develop consistent and uniform training outcomes. However, it was argued in this thesis that the STCW Code is too vague and therefore susceptible to individual interpretation in adopting learning and assessment processes towards competence development. The variability introduced by individual interpretation of the STCW creates the risk of seafarers graduating with CoCs but lacking the required competence for workplace operations. The seafaring industry sources its employees globally and cannot afford to operate under such risks. Based on the review of the selected excerpts, it was shown in this thesis that the STCW Code gives the

impression of being an input-based education system and not an outcome-based as it was originally intended to be. An input-based system may prove to be regressive for the seafaring industry due to its focus on curriculum and content coverage and not on the appropriateness, learning and assessment, or the attainment of the desired competence outcomes by the student. Graduates may be assessed as competent but lack the necessary attributes making it a point of risk for employers and governments relying on SET providers to deliver seafarers that meet the required standard.

The assessment methods currently in use by assessors in SET institutes are largely influenced by the STCW Code established by the IMO. However, traditional assessment methods failed to meet and fulfil all stakeholder expectations. This is especially the case for seafarer students who disengage with traditional assessments which require students to construct responses based on rote learning and analysis of information presented devoid of context. Hence, the focus of the IMO should be to convince the educators and assessors in the SET institutes to implement innovative assessment methods like authentic assessment under the current version of the STCW Code. To do so, evidence of the positive outcomes of such assessments on student learning is required. This study collected the required evidence on the impact of a new and innovative pedagogical practice (authentic assessment) on SET. This was conducted through a rigorous experimental design that compared seafarer students' academic achievement for traditional and authentic assessment. Findings of this research project found higher academic achievement of seafarer students in authentic assessment. Since, recommendations to the design and implementation of authentic assessment are essentially missing from the STCW Code, the findings of this research study may be used as a justification to add the recommendations to the code. The addition and subsequent implementation of authentic assessment in SET may bridge the resulting gaps between stakeholder expectations with the seafarer assessment process and the current practices.

Prior to this research study, the concept of validity and reliability in assessment was related to the interpretation of the student scores. However, it was argued in this thesis that validity and reliability are essential technical measures for evaluating the quality of authentic assessment; and the various aspects of validity and reliability need to be improved to achieve the intended outcomes of the assessment. While addressing different aspects of validity will identify and assess the content and essential underlying constructs of professional competence in different contextual scenarios; different aspects of reliability will assure consistency in performance. Overall, this will ensure a holistic approach to competence assessment at a standard expected in employment.

To ensure a holistic approach to competence assessment, a new and innovative hypothesized relationship was embedded into the evaluation of the concepts of validity

and reliability of authentic assessment in this thesis. The hypothesized relationship stated that 'if aspects of validity and reliability of authentic assessment are improved holistically, then assessment of seafarer education and the resulting evidence of student competence to perform workplace tasks can be significantly improved'. In this regard, a conceptual framework (AAFSET) to address the various aspects of validity and reliability during the different stages of assessment implementation was developed in this thesis. Using AAFSET to implement valid and reliable authentic assessment will change the focus of seafarer education from curriculum and content coverage to the legitimacy and consistency of learning and assessment. This is anticipated to lead to the attainment of the desired standard of competence required by the student more frequently and to reduce the variation in levels of competence.

A contribution of this thesis study was made through the collection of much needed empirical evidence on the impact of authentic assessment in seafarer education since similar research has not previously been conducted. On the basis of the comparison between authentic and traditional assessment scores, the findings of this research confirmed that students assessed authentically had significantly higher scores resulting in higher academic achievement. This finding indicated that students' academic achievement will be improved if they focus on the assimilation, critical analysis, and integration of information presented through a real-world context instead of memorising information and rote learning. The findings of this research also indicated that students' academic achievement will be improved if they are provided with feedback that may be used in recognizing gaps in their knowledge and skills; and then at least one opportunity to attempt a similar task before the judgement on their competence is made.

Factors of authentic assessment (task, context, transparency of criteria, etc.) that correlated significantly to higher academic achievement (measured using scores obtained in the assessment tasks) was also investigated in this research study. Findings derived through factor analysis confirmed that the factor transparency of criteria (confidence level 95%) was a significant predictor of student scores. The performance criteria was made transparent in this research study through the use of assessment rubrics provided to the students at the beginning of the learning period. Accessibility to the rubrics beforehand made the students fully aware of the learning expectations and reduced uncertainty and anxiety with the assessment practices, thus creating higher engagement with learning. High student engagement promotes deep understanding of learning content and motivation to master skills and knowledge.

Accessibility to assessment rubrics that detailed performance criteria, engaged students in meaningful reflection and self-assessment of their task performance, while being actively involved in the process of knowledge construction through the critical analysis of information presented through a demonstration of a case-study in a real-world

context. This may have promoted a deep connectedness of learning with real-world applications. Hence, this research uncovered significant factors of assessment which, if included, in the design of the assessment will not only guide authentically assessed students towards higher academic achievement but make them more competent learners and professionals. This study not only enabled the researcher to find responses to the questions raised and hypotheses tested but also opened up new avenues for further discovery and reflection into using alternative approaches that are cost-effective for evaluating seafarer competence while integrating standardized assessments.

7. APPENDED PAPERS

Chapters 1 to 6 were designed to demonstrate that the papers included in this thesis constitute essential parts of a coherent and integrated whole, and that they contribute to the overall research work. The final chapter (Chapter 7) reproduces the seven manuscripts published or submitted in the framework of the doctoral project. The seven papers have not been rewritten for this thesis. There are, therefore, unavoidable repetitions, especially among the papers as well as between the papers and the thesis chapters. However, in order to ensure the consistency of format, the seven papers have been reformatted and the references have been included in the unique list of references at the end of this thesis. With a view to better differentiating the actual papers from the remainder of the thesis, a different format (font, font-size, etc.) has been used for the overall presentation of the papers, i.e. for the headlines, the content, the illustrations, etc. Table 7.1 recaps the characteristics of the papers reproduced in the remainder of this chapter.

Table 7.1: Publication recap

Paper	Nature	Title	Publication channel	Full-length, double blind review	Status
I	Conceptual	Reviewing seafarer assessment methods to determine the need for authentic assessment	*Australian Journal of Maritime & Ocean Affairs*	Yes	Published (2014)
II	Conceptual	On a lookout beyond STCW: Seeking standards and context for the authentic assessment of seafarers	*IAMU AGA 15 Looking Ahead: Innovation in Maritime Education, Training and Research*	Yes	Published (2014)
III	Conceptual	Using authentic assessment to enhance seafarer student engagement and their ability to transfer learning	*IAMU AGA 2015 International Association of Maritime Universities: 16th IAMU Annual General Assembly*	Yes	Published (2015)
IV	Conceptual/ Research	Authentic assessment in seafarer education: using literature review to investigate its validity and reliability through rubrics	*WMU Journal of Maritime Affairs*	Yes	Published (2016)
V	Conceptual/ Research	Improving the validity and reliability of authentic assessment in seafarer education and training: A conceptual and practical framework to enhance resulting assessment outcomes	*WMU Journal of Maritime Affairs*	Yes	Published (2017)
VI	Research	Authentic versus traditional assessment: An empirical study investigating the difference in seafarer students' academic achievement	*Accepted for publishing in Journal of Navigation*	Yes	Under review
VII	Research	Investigating the correlation between students' perception of authenticity in assessment and their academic achievement in the associated assessment tasks	*Under review in journal*	Yes	Under review

7.1. Paper I

REVIEWING SEAFARER ASSESSMENT METHODS TO DETERMINE THE NEED FOR AUTHENTIC ASSESSMENT

Samrat Ghosh
Australian Maritime College, University of Tasmania
Marcus Bowles
Australian Maritime College, University of Tasmania
Dev Ranmuthugala
Australian Maritime College, University of Tasmania
Ben Brooks
Australian Maritime College, University of Tasmania

Abstract

The Standards of Training, Certification and Watchkeeping (STCW) Convention set global, minimum standards of competence for seafarers. Maritime Education and Training institutes (METs) are responsible for ensuring assessment processes not only assure attainment of STCW outcomes but produce competent graduates that meet the expectations of the core stakeholders such as regulators and employers. However, review of literature in this area suggests that some of the current assessment methods employed by METs are largely failing on both accounts. This paper argues that STCW falls short in its ability to provide appropriate standards and looks at the need of authentic assessment in seafarer training. A brief review of authentic assessment presented herein highlights characteristics that may improve the shortcomings of current assessment methods and STCW. It aims to propose authentic assessment as a way to elevate the collection of evidence of a seafarer's competence using methods that promote student engagement and transfer of competence in different contextual scenarios.

1. Introduction

Accidents in the seafaring industry can have catastrophic effects leading to loss of lives and large oil spills that damage the marine environment. Analysis of accidents have often revealed that the lack of sufficient competence to operate a ship and its systems have often contributed to such accidents (Pecota and Buckley, 2009). The International Maritime Organization (IMO) introduced the Standards of Training, Certification and Watchkeeping (STCW) Convention in 1978 (referred to as STCW'78) with one of its objectives aimed at reducing human error due to the lack of competence. The IMO intended to achieve this by establishing global, minimum standards of competence that provide guidelines to Maritime Education and Training institutes

(METs) for assessing seafarers via a standardized system of competence assessment based on which the latter can be granted a Certificate of Competence (CoC). The CoC opens job opportunities for seafarer students within seamless career pathways that, based on competence, can take them from entry level to ship's Captain and beyond. National maritime regulatory bodies use the STCW approved CoC as a basis for regulation of the workforce on ships registered under their flags. Employers consider the CoC to be an evidence of a seafarer's competence to perform at the workplace and a basis for them to recruit, reward and train their employees. Thus, employers, maritime regulatory bodies, METs and the seafarers become the key stakeholders for this paper. The assessment process leading to the issuance of the CoC, becomes a key component to fulfil the expectations associated with it for these key stakeholders. However, review of literature in this area provides arguments to support current assessment methods are largely failing to fulfil expectations, in particular employer expectations. This paper highlights stakeholder expectations from the assessment and certification process of the seafarers. It makes a theoretical contribution by identifying the gaps between current and expected seafarer training outcomes resulting from some of the current assessment methods in use by global METs and those promoted by STCW.

2. Background

Assessment is a significant component of education along with learning and teaching. It can be conducted at various stages of the learning cycle with its outcomes providing feedback about a student's progress and achievements; the effectiveness of the teaching and instruction methods; and the course outcomes while fulfilling the overall goal of improving student learning [University of Tasmania (UTAS), 2011]. In the context of professional education and training, such as seafaring, assessment also provides feedback about the achievement of professional standards by students that are essential for the workplace. Attainment of such standards provide evidence of an ability to combine knowledge, skills, values and attitudes into behaviours (Aranda and Yates, 2009) required to perform in the real world which defines the student's professional competence (Rychen, 2004). Professional competence requires essential cognitive abilities of recalling information (knowledge) and applying it (skills) based on analytical and critical thinking (Nusche, 2008). Underlying are the values and attitudes that are non-cognitive skills that respectively shape the principles of thought and prompt responses based on the reflection of those thoughts (Moore and Asay, 2013). Seafarers can acquire professional competence in workplace settings on board ships and in METs under academic guidance as the training structure of a seafarer is divided into college and sea-based training alternately.

In the seafaring industry, global, minimum competence standards were established through STCW by the IMO. The IMO is the over-arching regulatory body which together with national maritime regulators enforce the requirements of STCW on maritime nations (or States) that have ratified the convention. The national regulator is usually known as the 'Flag State' (also known as Flag State Control or FSC) which can be an administration or the government of the State under which ships can be registered, e.g. Australian Maritime Safety Authority (AMSA) in Australia. The FSC becomes 'Port State Control' (PSC) when ships of other registry call at

their ports. The FSCs and PSCs of nations that have ratified STCW ensure compliance through inspections and surveys for all ships irrespective of registry when they visit ports of signatory nations. Countries fully complying with the provisions of STCW and its training and education practices are listed as the 'white list' of nations in the maritime industry. IMO has authorized the national maritime regulators to grant approval to METs for conducting STCW approved courses if, after a thorough inspection of their facilities, they are found to be in full compliance with the provisions of STCW. Figure 1 describes the roles of the key stakeholders in the implementation of STCW in a nutshell.

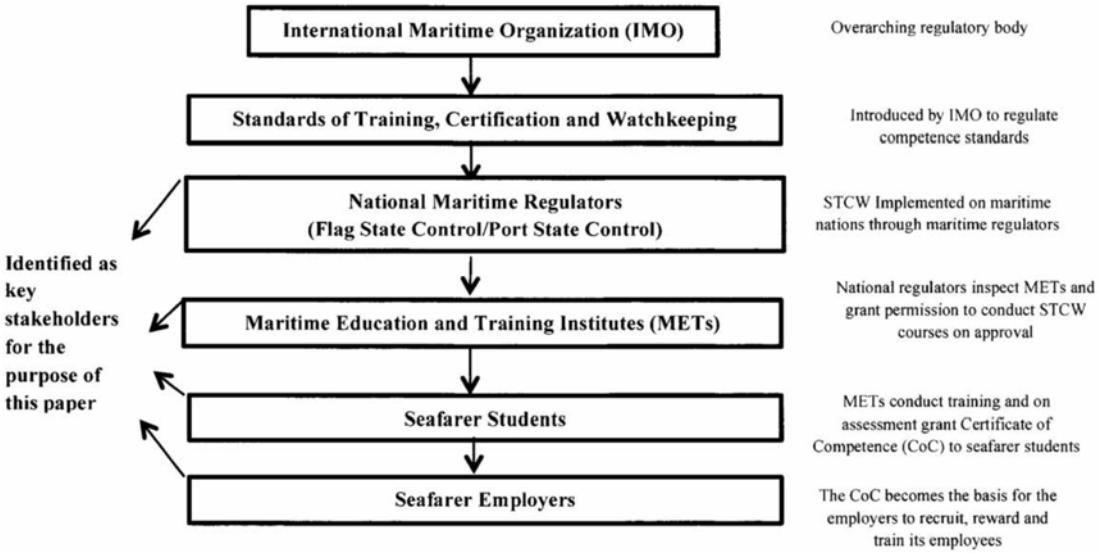


Fig.1 The role of the key stakeholders in the implementation of STCW

The STCW applies to seafarers who are working or intending to work on commercial vessels on domestic (coastal) or international voyages but not to those serving on naval vessels, government-owned or operated vessels engaged in non-commercial service, fishing vessels, pleasure yachts not engaged in trade and wooden ships of primitive build (STCW, 2011). It provides guidance to stakeholders about the competence that needs to be developed to safely operate ships, with an aim to create consistent and uniform competence standards in a global industry. STCW has been revised twice (1995 and 2010) since its conception in 1978. Currently, the revised STCW is referred to as STCW'95 including 2010 Manila Amendments. Figure 2 and Table 1 describes the standards of competence and assessment as laid out in STCW'95 (after its last revision in 2010).

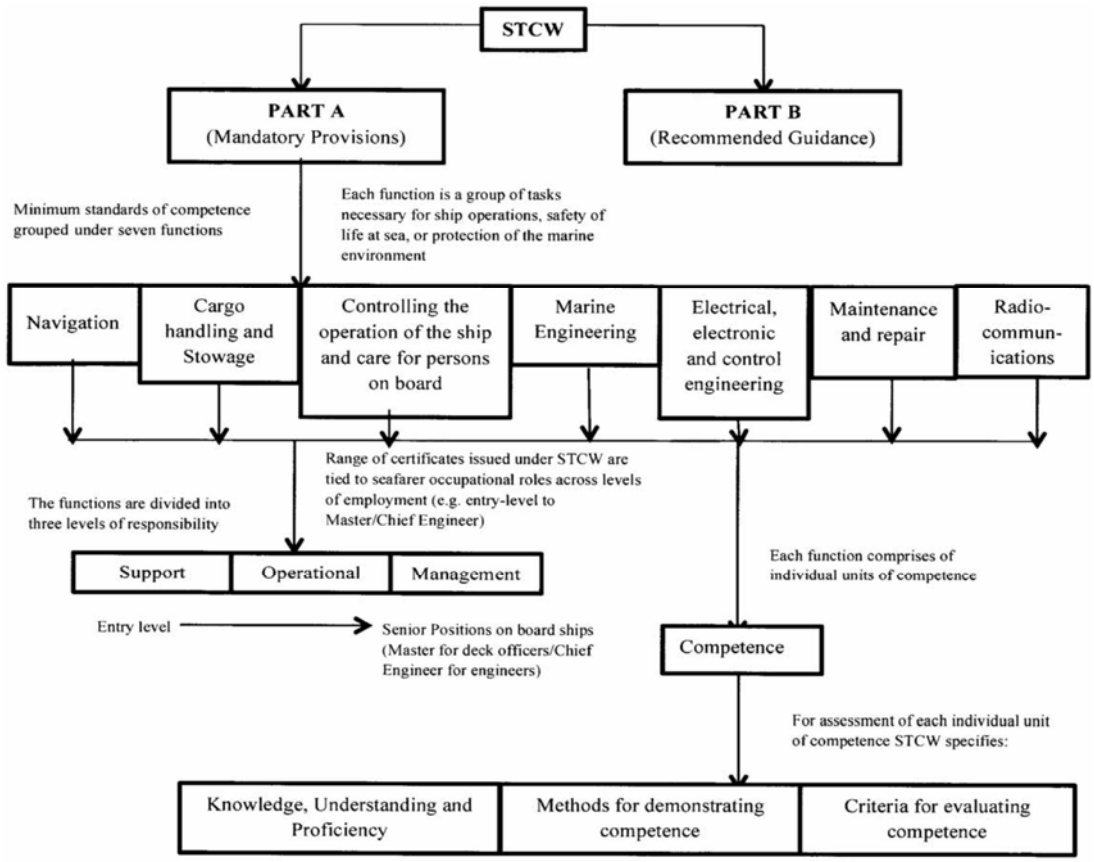


Fig.2: Standards of Competence and Assessment in STCW'95 (after 2010 revision)

Competence	Knowledge, understanding and proficiency	Methods for demonstrating competence	Criteria for evaluating competence
Carriage of dangerous goods	International regulations, standards, codes and recommendations on the carriage of dangerous cargoes, including the International Maritime Dangerous Goods (IMDG) Code and the International Maritime Solid Bulk Cargoes (IMSBC) Code Carriage of dangerous, hazardous and harmful cargoes; precautions during loading and unloading and care during the voyage	Examination and assessment of evidence obtained from one or more of the following: .1 approved in-service experience .2 approved simulator training, where appropriate .3 approved specialist training	Planned distribution of cargo is based on reliable information and is in accordance with established guidelines and legislative requirements Information on dangers, hazards and special requirements is recorded in a format suitable for easy reference in the event of an incident

Table 1. STCW minimum standards of competence for assessment: Example – Function: Cargo handling and stowage at the management level (STCW, 2011).

Typically working under the oversight of the maritime regulator, it is the responsibility of the METs to assess and collect evidence of a students' ability to apply knowledge learnt in

classroom and during seatime (service on board a ship) to practical skills at appropriate levels of responsibility (e.g. support, operational and management), in accordance with STCW. However, review of literature (Sampson, Gekara and Bloor, 2011; Emad and Roth, 2007) in this area suggests that the needs of some of the key stakeholders discussed in this paper are not being fully met by METs. This paper argues that STCW falls short as the global 'currency' promoting well defined and appropriate assessment methods and in providing performance standards appropriate for stakeholder expectations.

3. STCW Fails Expectations and Causes Concerns

Prior to STCW'78, countries established their own standards. However, STCW'78 did not prove to be as effective as expected due to criticisms from stakeholders that complained of vague and unclear standards left to the individual interpretations by maritime nations [International Maritime Organization (IMO), 2013b]. As a result, STCW was revised with amendments in 1995 (referred to as STCW'95) to address previous concerns and improve upon the training mandate to make it outcome based requiring seafarer students to demonstrate their competence by performing tasks which resembled workplace duties (Emad and Roth, 2008). With the 2010 manila amendments, STCW intended to not only lay the competence requirements for a seafarer but also make parties to the convention accountable (Leca Da Veiga, 2001) [to] the various stakeholders and each other for the quality of their certification process.

However, STCW'95 did not fully eliminate the vagueness in assessment standards as it specified methods to demonstrate competence but did not provide specific methodologies, leaving it to the discretion of the assessor (Robson, 2007). For example, from Table 1 it is evident that STCW specifies methods (simulator, specialist training, etc.) to demonstrate competence but does not provide clear and detailed guidelines on how to use these methods to collect evidence of competence. For example, how sophisticated and advanced should the simulators be to reflect STCW standards? The STCW only provides recommended performance standards for non-mandatory types of simulators.

Even after the last revision of STCW in 2010, the vagueness in STCW continues to leave too much room for interpretation by METs, who are using varying combinations of assessments (Bhardwaj, 2009; Drown, Mercer, Jeffery and Cross, 2010; Kean, Matthews, Meadows and Stone, 2011) for students to demonstrate the performance standards in STCW. Performance standards should ideally communicate performance expectations from workplace duties, encompassing not only the technical skills but also the underpinning skills and knowledge. However, table 1 provides us with an example on how STCW fails to recognise all necessary technical and underpinning skills or units of competence. For example, planned distribution of cargo and recording information are not the only skills required for carrying dangerous goods. It should also identify essential underpinning skills such as problem identification and solving if there are any unexpected occurrences with its carriage. The METs complying strictly with STCW will assess seafarers in accordance to the inadequate performance standards

producing graduate seafarers lacking workplace skills. This is a major concern for seafarer employers.

4. Expectations and Concerns of Seafarer Employers

Seafarers are entrusted with ships worth millions of dollars and lives of their colleagues that are priceless. Seafarer employers' need reliable indication or contextual evidence of their employees' competence so that gaps in knowledge and skills can be identified and filled with additional training, if required. Assessments that do not provide contextual evidence may leave employers clueless as to what should the additional training should focus on. Costs of additional training are often bore by the employer (Hanzu-Pazara and Arsenie, 2007). Although employers have training obligations for preparing their employees for specific types of vessels, costs borne for aimless training should be avoided as it can cause a significant impact on the employers' budgets. Many employers already feel reluctant to spend capital on employee training due to the risk of them being poached by other companies offering higher salaries (DEEWR, 2010).

The International Safety Management (ISM) Code developed for the safe operation of ships clearly states that it is the responsibility of the seafarer employers to ensure their employees are competent to work on board ships (IMO, 2002). The IMO authorises national regulators to investigate seafarers' competence through inspections and surveys, amongst many other regulatory requirements, to identify and deter substandard ships from operating (AMSA, 2011). Ships can be detained, and registers cancelled if serious deficiencies are found in an operators' ability to perform workplace tasks safely (Department of Infrastructure and Transport, 2012). Due to a global shortage of seafarers, training period of trainee cadets have been reduced and young officers with a reduced sea experience are being promoted (Listewnik, 2009) to fill up the higher ranks of responsible officers on ships, on obtaining the CoC for the appropriate level of responsibility. Hence, many progressive employers are investing large amounts of capital for training seafarers (Sadjadi and Perkins, 2010) in METs expecting the certification process to result in graduates that have high standards of competence.

However, a study by (Sampson, 2011) revealed that many seafarer employers are aware of the varying standards of the assessments in global METs and regard the resulting CoC as crating a 'lowest common denominator' standard of achievement that provides an unreliable proof of competence for every workplace. The study comprised of a series of seventeen interviews with employers in the UK, Philippines and Singapore, that recruited seafarers on vessels involved in international voyages. The interviewees were fleet personnel managers from both owner operator companies and of ship management companies, involved in the employment of seafarers largely from countries like India, Myanmar, and Philippines and from Eastern Europe.

The study revealed that employers were unhappy with some of the current assessment methods that assessed a limited range of job specific skills, in settings that provide insufficient cues to the students on how the competence acquired in classrooms can be used in different contexts. For example, official investigations and analysis of marine accidents have revealed that

seafarers assessed as competent in the use of particular skills in a context have failed to apply them in another.[23]

The practice of assessing a limited range of skills can also curtail the development of a holistic portfolio all the necessary skills required for supporting workplace performance at a particular level of responsibility (Cox, 2009). This can be elaborated using an extract from STCW, for the function of navigation at the operational level.

Competence	Knowledge, understanding and proficiency	Methods for demonstrating competence	Criteria for evaluating competence
Respond to a distress signal at sea	Search and Rescue Knowledge of the contents of the International Aeronautical and Maritime Search and Rescue (IAMSAR) Manual	Examination and assessment of evidence obtained from practical instruction or approved simulator training, where appropriate	The distress or emergency signal is immediately recognized Contingency plans and instructions in standing orders are implemented and complied with

Table 2: Extract from STCW for the function of navigation at the operational level (STCW, 2011)

The extract shows that competence for carrying out an effective search and rescue at the operational level can be assessed without identifying and testing for essential underpinning skills, such as ability to report information without distortion or filtering; ability to prioritize according to urgency of the situation, etc. It may be argued that STCW does not specify the supporting skills in all its units of competence, but can the seafaring industry afford to hide behind this oversight? Employers expect underlying competencies to be assessed along with the technical skills (Cross, 2007; Cox, 2009). Competencies are skills distinct from technical and work-related skills which when use singularly or in various combinations integrally with the technical skills, support the performance which defines competence (Teodorescu, 2006; Rutherford, 1995). Investigation of shipping accidents have often indicated seafarers lacking such skills (Hetherington, Flin and Mearns, 2006; Devitt and Holford, 2010). The assessments designed should ensure that such underlying competencies are identified and suitably assessed in a unit of competence to develop the skills in the seafarers.

The extract above also highlights that STCW encourages assessment for individual competence units and not a holistic approach to assessment. Students should be able to integrate the competence gained from individual units in STCW, for a successful performance reflecting workplace standards. Using the example from table 2, the competence to respond to distress signals at sea should be simultaneously assessed with the ability to carry out routine duties of navigation and vessels' passage monitoring, as may be required at the workplace.

5. Expectations and Concerns of Seafarer Students

Successful performance in the workplace or during assessments in METs would require skills developed in a particular context to be transferred in different contexts and varied scenarios. However, the transfer of skills is affected by the context in which it was developed (Leberman, Mcdonald and Doyle, 2006). There are skills defined by some authors (Clanchy and Ballard, 1995; Nusche, 2008) as transferable (or generic) skills that are not tied to particular contexts and are directly transferred to different scenarios. For example, a seafarer who has developed oral and written communication skills or the ability to plot a ship's position using GPS data on hydrographic charts should be able to do so irrespective of the context or situation. On the other hand, there are non-transferable (domain specific) skills that are defined by some authors as skills tied to particular contexts, with students requiring learning how to use them in different scenarios (Clanchy and Ballard, 1995; Nusche, 2008). For example, a seafarer with skills to manoeuvre a fifty-metre ship may not be able to do so when the ship's length increases to three hundred metres. The assessment process can play a key role in acting as a "pit stop" where students and assessors can reflect on the application of their skills (Curry, Caplan and Knuppel, 1994) in a particular context and identify additional training requirements for different contexts.

The outcome of seafarer assessments should be to inculcate such skills in students in order to allow them to make appropriate decisions in routine as well as unperceived situations and adapt to the diverse shipboard environment (Prasad, 2011). To do so, students should be assessed via tasks in different contextual scenarios that resemble workplace scenarios that produce sufficient and reliable evidence of a seafarer's competence (Gonczi, Hager and Athanasou, 1993). Although it may not be possible to recreate exact shipboard situations due to the complexity in structure and design, assessments designed to closely resemble workplace situations may provide cues for transfer of competence for the student in the real world (Mcmullen and Braithwaite, 2005). This indicates that context similarity may be very relevant in seafarer assessment.

Lack of contextual similarity also raises the questions of validity where students start to question the relevance of the assessments and the competence it purports to assess. For example, student interviews during the course review of Deck officers carried out in the Australian Maritime College (AMC) in 2011 revealed that seafarer students could not relate to traditional pen and paper testing when it came to assessing competence for performing workplace tasks. They showed a preference for assessments that are contextually similar to challenges found at workplace, in order to relate classroom learning to professional practice [Australian Maritime Organization (AMC), 2011].

An ethnographic case study involving 16 students in a Canadian maritime institute revealed that students were aware of the fact that traditional exams comprised mainly of the questions which were drawn from a question bank (Emad and Roth, 2007). Over time they could predict the range of questions and prepare accordingly. Such assessment methods that do not provide much scope for innovation in design; can encourage memorizing to pass examinations. Providing memorized answers do not reflect the actual competence of the student to perform the same task on board a vessel (Lewarn, 2002). Ability to memorize is a lower level cognition

which is not sufficient for performing in workplaces such as ships, where a higher level of cognition is required to assimilate, analyse and structure information for decision making and problem solving (Wiggins, 1990).

The STCW recommends that a seafarer's competence should not only be determined by their ability to integrate knowledge and skills in routine contexts but also by their ability to operate in unique and constantly changing conditions that may require critical thinking and higher order cognitive skills (Walczak, 1999). Seafarers who are trained to rely on memory and not to assimilate and analyse the available information to deal with routine or novel situations in the context of the work environment, may suffer from memory failure leading to human errors (Prasad, 2011). For example, a seafarer may have been certified as competent through assessment based on memorized answers in an oral examination. However, memory lapses may lead to an unintentional skill and knowledge-based errors at the workplace. For example, a recent report by the IMO, showed the distribution of detailed causes for human errors, technical failures and external factors, based on statistics from the Norwegian Maritime Authority (Figure 3). Poor judgement (29%), inadequate competence (12%) and lack of knowledge, skill, and training (3%) were some of the factors highlighted as causes of human error. Competence acquired through analysing and assimilating information train professional students to select the correct course of action based on gathered evidence and not purely on memory.

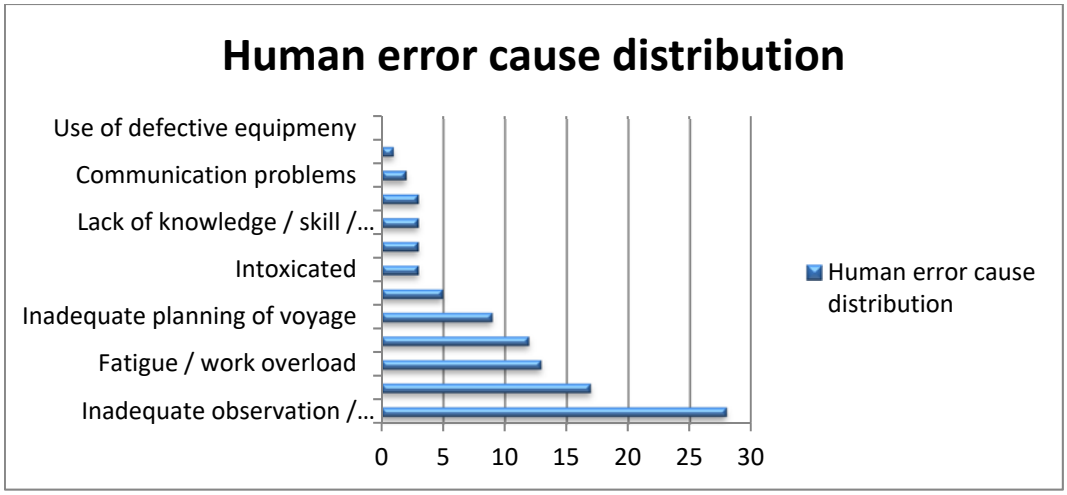


Figure 3: Human error causes of navigational accidents (International Maritime Organization, 2013a)

6. Expectations and Concerns of Maritime Regulators

As discussed before in this paper, STCW falls short in providing specific methodologies for assessment. For example, STCW does not provide performance standards for simulators to be used for assessment. The recommendations provided are non-mandatory which maritime nations do not have to necessarily follow. Many countries, including some of the 'white list" nations, allow METs to operate with training and assessment regimes that barely meet the

minimum compliance of STCW. For example, a study showed that Philippines is a 'white list' nation that has allowed sub-standard METs to operate in spite of falling short of STCW expectations. The ethnographic study focused on Singapore, United Kingdom and Philippines and comprised of thirty in-depth interviews with company managers, college lecturers and trainers, union officials and a member of the IMO (Sampson, 2004).

Maritime nations often have excessive competition amongst its METs. The METs need to provide economical and affordable training to attract more students (Bloor and Sampson, 2009) due to which they may not invest in costly simulators and other training/ assessment facilities. Another view of the same issue can be seen from the eyes of the METs from developing or low-cost nations who intend to fully comply with STCW. However, they may be unable to afford costly training and assessment facilities like simulators and seek support from other stakeholders, such as the national government. In many instances, such support might not be available to METs from its stakeholders (Baylon and Santos, 2011).

The concerns arise from the fact that many seafarer employers source their employees from developing or low-cost nations (Department of Infrastructure and Transport, 2012) for reasons such as reduced labour costs, competitive pressures, shortage of seafarers in developed nations, etc. (Theotokas and Progoulaki, 2007). Bare minimum compliance of STCW training and assessment standards or absence of facilities may result in producing seafarers lacking the required knowledge and skills to deal with changing contexts at the workplace. The reputation of flag states may seriously be damaged in the international community (Department of Infrastructure and Transport, 2012) if the same seafarers sailing under their ship registers with STCW approved CoCs, become contributory factor to accidents.

One of the ways regulators measure their own performance in implementing various IMO requirements for ship operations, is through the reduction of deficiencies per inspection, of foreign flag ships operating in their waters (AMSA, 2013a). This measure of performance may be affected if seafarers having sub-standard competence increase deficiencies in ship operations. Sub-standard competence can be associated with seafarers graduating from METs with inadequate facilities as well as from METs having good training facilities but utilising inappropriate assessment methods. This falls back on the shoulders of Flag States who granted METs approval to conduct STCW courses in the first place.

Regulators often issue Certificate of Endorsement (CoE) to seafarers holding overseas qualifications to fulfil shortage of skills in their respective nation (AMSA, 2011). Such issuance may at times involve assessment of seafarers. For example, in Australia, seafarers holding overseas certificates must pass an AMSA oral examination to obtain an AMSA CoC for the equivalent level of responsibility (AMSA, 2013b). In such cases, it becomes essential that the assessment and resulting certification process reflect best practice and contemporary needs of the industry to identify skill shortages in the seafarer.

7. Concerns and expectations of Maritime Education and Training Institutes (METs)

The METs are expected to strictly comply with STCW requirements by their respective national maritime regulators. But how can these institutes design assessment methods when STCW fails to provide unambiguous and clear terms for assessment of many units of competence? For example, Table 3 highlights words (promptly, minimize, etc.) from an extract in STCW that do not provide clear terms of measurement. The words fail to provide a benchmark against which METs can assess how quickly, in terms of speed, accuracy or completeness the candidate should demonstrate the desired level of performance (Rutherford, 1995). In the absence of a well-defined meaning, these words may be differently interpreted, leading to subjective assessment.

Competence	Knowledge, understanding and proficiency	Methods for demonstrating competence	Criteria for evaluating competence
Respond to navigational emergencies	Precautions when beaching a ship Action to be taken if grounding is imminent, and after grounding Refloating a grounded ship with and without assistance Action to be taken if collision is imminent and following a collision or impairment of the watertight integrity of the hull by any cause Assessment of damage…….	Examination and assessment of evidence obtained from practical instruction, in-service experience and practical drills in emergency procedures	The type and scale of any problem is promptly identified, and decisions and actions minimize the effects of any malfunction of the ship's systems Communications are effective and comply with established procedures Decisions and actions maximize safety of persons on board

Table 3: Extract from STCW for the function of Navigation at the management level (STCW, 2011)

An issue of concern may also be the absence of levels in the STCW competence standards, i.e. the tasks are not based on a framework which provides students with a pathway to progress from simple to complex levels of the task (Rutherford, 1995). The current standards do not identify procedures for METs to design assessment methods that would allow seafarer students to gain expertise in performing the task in stages moving from their current level of competence to the required level of expertise (Benner, 2001).

The METs do not have the authority or power to revise assessment standards in STCW. They can also not be expected to meet and fulfil all stakeholder expectations. Barely meeting the existing standards or the regulator's expectations will not satisfy the expectations of the employers or optimise employability or competence of seafarer students. However, the METs can design assessment methods or tasks that can identify the gap between the current STCW

standards and standards actually required for the workplace tasks. The gap may be recognized as underpinning competencies or improved performance standards, which may be added to the STCW for improvement in seafarer assessment methods.

8. The Way Forward

Evidence from past literature presented herein suggests that current assessment methods in seafarer training, to a large extent, are not fulfilling the expectations and addressing the concerns of the key stakeholders. Table 4 summarises the key concerns and expectations discussed in this paper.

C&E No.	Concerns and Expectations (C&E)
1.	Varying standards of assessment in METs due to vagueness in STCW
2.	Insufficient cues for transfer of competence during assessments for different contextual scenarios
3.	Assessment of technical (job specific) skills only and not the underlying competencies
4.	Overlooking the assessment of an ability to integrate competence acquired in individual tasks
5.	Lack of contextual similarity to workplace in assessment tasks
6.	Lack of validity (assessments not testing what they should) and reliability (lack of multiple indicators of competence)
7.	Assessments focussing on memorizing knowledge and not on critical thinking and higher order skills/knowledge
8.	Lack of levels in STCW competence standards to provide students with a pathway to gradually practice workplace tasks to reach required expertise

Table 4: Summary of concerns and expectations of key stakeholders

The concerns and expectations can be divided into two broad categories in Table 5.

Shortfalls in STCW	Shortfalls in current assessment practices
C&E No.1; No. 8	C&E No.2; No.3; No.4; No.5; No.6; No.7

Table 5: Distribution of Concerns and Expectations in categories

Table 5 highlights the need for an assessment method that has suitable characteristics to address these concerns and expectations. While only IMO can document official changes in the STCW, assessors in IMO can use assessment methods that can improve the resulting competence to standards beyond STCW. C&E Nos. 2, 5 and 6 need to be addressed by assessments which are characterised by tasks contextual to the workplace situations (Herrington and Herrington, 2006). Contextual tasks may be found meaningful by the seafarer students due to its strong figurative context and fidelity to the situations that they may find themselves in the professional world. Such assessment methods have often been defined as authentic assessments by some authors (Wiggins, 1989; Gulikers, Bastiaens and Kirschner,

2004). Contextualised authentic tasks may not recreate all the conditions of the workplace but will replicate the complexities and challenges (Frey, Schmitt and Allen, 2012) in different scenarios which will require students to integrate a range of competencies for problem-solving and decision-making as in the real world, providing multiple indicators of competence. Seafarer students may then find learning more engaging as they acquire professional skills, thus enhancing transfer of competence (Tennant, 1999).

Authentic assessments may be designed to be a continuous process integral with the learning and teaching, which will allow students to practice skills till they reach the required level of competence (Benner, 2001). It will also allow students to frequently reflect (Herrington, Reeves and Oliver, 2010) on their work to recognize gaps in their knowledge and grasp cues for enhancing transfer of context free transferable skills and domain specific non-transferable skills (Wiggins, 1993). Authentic assessment promotes providing students access to the resources that would available to them in the outside world (Gulikers, Bastiaens and Kirschner, 2004) to move the students beyond memorizing and make them focus on developing an in-depth understanding by assembling and interpreting information, formulating ideas, critiquing, integrating knowledge and holistic application of skills (Archbald and Newmann, 1988) for higher order cognition (Tanner, 2001). Students also have access to the performance criterion (reflecting workplace requirements) beforehand for them to aim for the same level of ability in performing assessment tasks. This will not only lead to standardized competence levels across the range of tasks (Lund, 1997) but also enable assessors to improve upon existing STCW standards and provide multiple indicators and contextual evidence (Linn, Baker and Dunbar, 1991) of a seafarers' competence.

Table 6 summarises how the characteristics of authentic assessment may address the concerns and expectations raised in the paper.

Contextual Similarity	Integration of competence/ competencies	Opportunity to practice skills	Student Reflection	Access to professional resources	Performance criterion and standards that support learning
C&E No.2	C&E No.3	C&E No.2	C&E No.2	C&E No.2	C&E No.1
C&E No.3	C&E No.4	C&E No.4	C&E No.3	C&E No.4	C&E No.3
C&E No.4		C&E No.7	C&E No.7	C&E No.7	C&E No.6
C&E No.5		C&E No.8			C&E No.9
C&E No.6					
C&E No.7					

Table 6: Characteristics of Authentic Assessment and the concerns and expectations (C&E) they will address

9. Conclusion

Student assessment should not only be a formality conducted as an educational practice but a process that should be carefully planned and implemented to serve the focal goal of improving student learning. In the context of seafaring education, assessment of seafarers forms the basis on which they can be granted a CoC for particular roles and levels of responsibility, providing them with a licence to operate ships and its' equipment. Ships are "floating structures" at sea which not only carry human resource, but in many cases, different volumes of cargo, oil and other entities. Past records indicate that accidents involving ships have often resulted in loss of lives and damage to the marine environment, affecting not only the stakeholders involved in the shipping industry but sometimes the entire community. Analysis of shipping accidents has proved that a major percentage of these accidents are caused by operational mistakes and errors by the ships' crew.

Seafaring industry will always be at risk of major operational errors if competence levels of seafarer operators are not accurately and adequately assessed before issuing them with the CoC. The assessment methods currently in use by assessors in METs are largely influenced by the STCW established by the IMO. The outcome of the assessments, in the form of a CoC, provides justification for its holder to seek job opportunities; for the employers to recruit, reward and train; and for the maritime regulators to form a workforce that comprises of professionals with standardized competence levels.

However, a review of current seafarer assessment methods presented herein highlights the flaws in the current assessment methods and shows a need for improvement in order to bridge the resulting gap between current and required stakeholder expectations. The METs are powerless in revising STCW standards on their own and cannot be fully blamed for interpreting and adopting vague assessment methods promoted by STCW. Assessment methods for seafarer training need to address major concerns as identified in the paper. The key issues include, assessment methods failing to develop skills that enable seafarer students to transfer their competence from METs to workplace contexts; assessment methods lacking relevance and meaning to students disengaging them from the learning process; and varying standards of assessments in global METs which are likely producing seafarer graduates with inconsistent and sometimes inadequate competence. The use of authentic assessment in seafarer training may be one of the methods to address these issues. A brief description of authentic assessment mentioned in the paper suggests that it may contain the necessary characteristics which if implemented in seafarer training, may assist in resolving issues of student engagement, skills transfer and inconsistent competence standards. It will also identify essential underpinning skills and improve upon existing STCW performance standards to match workplace requirements. Future research aims to validate these theoretical arguments resulting in outcomes and evidence which may be presented to the stakeholders for judgement and subsequent implementation.

7.2. Paper II

ON A LOOKOUT BEYOND STCW: SEEKING STANDARDS AND CONTEXT FOR THE AUTHENTIC ASSESSMENT OF SEAFARERS

Samrat Ghosh
Australian Maritime College, University of Tasmania
Marcus Bowles
Australian Maritime College, University of Tasmania
Dev Ranmuthugala
Australian Maritime College, University of Tasmania
Ben Brooks
Australian Maritime College, University of Tasmania

Abstract

The Standards of Training, Certification and Watchkeeping Convention (STCW) amendments in 1995 intended to improve the knowledge-based training mandate established in STCW'78 by making it outcome-based. This required seafarer students undertake competence assessment (or outcome of training received) to demonstrate their capacity to perform tasks listed in the STCW Code. This necessitates that students direct their learning efforts to the attainment of clearly stated expectations that, typically, are represented by learning outcomes based on the STCW competencies. Maritime Education and Training (MET) providers working under the directives of the National Maritime Regulators interpret the STCW requirements to develop the seafarer training curriculum and the resulting learning outcomes, to assure that students attain the minimum standards of competence established by the STCW. This paper will review and argue that different ideas as to 'outcomes' has been confusing the interpretation of STCW and, therefore, how seafarer students are being assessed. Critically, a review of specific excerpts from the STCW Code will be used to show that the Code largely fails to provide a 'standard' that can assure assessment of seafarers to one of the most critical outcomes: the performance expected at a level of work in the industry. A short review of the inherent characteristics of authentic assessment is provided in justification of its use as an alternate and optimal solution to improve current assessment practices and respond to stakeholder needs. The paper will point to an evidence-based way forward where future research will empirically investigate how authentic assessment can improve the STCW and the resulting training outcomes.

Keywords – STCW, Outcomes, Criteria, Standards, Context, Authentic Assessment

1. Introduction

The International Maritime Organization (IMO) established the Standards of Training, Certification and Watchkeeping (STCW) Convention in 1978 (referred to as STCW'78) to provide global, minimum standards of competence for seafarers. Prior to STCW'78, individual countries established their own standards. However, STCW'78 did not prove to be as effective as expected due to criticisms from stakeholders that complained of vague and unclear standards left to the individual interpretations by maritime nations [International Maritime Organization (IMO), 2013b], which posed the risk of variation in the standards of competence development amongst international seafarers. To address these concerns and improve upon the training mandate, the STCW Code was revised with significant amendments in 1995 (referred to as STCW'95). Through the 1995 amendments, IMO intended to fundamentally improve the training mandate by making it outcome-based. This would require seafarers to demonstrate their competence in the tasks outlined in the STCW Code rather than just show they had acquired knowledge (as in STCW'78). Over the years STCW has been updated with various amendments (1997, 1998, 2004, 2006, Manila amendments 2010) to provide training and assessment guidelines to Maritime Education and Training (MET) providers and other stakeholders with an interest in developing the competence seafarers require at the workplace.

The STCW Convention developed the STCW Code which provides guidelines on what the seafarer student should know and demonstrate before being awarded with the Certificate of Competence (CoC). The CoC opens job opportunities and based on competence, becomes the basis for their recruitment, reward and promotions. The Code promotes specific assessment methods to collate evidence of demonstrated competence for the tasks listed in it. However, both competence demonstration and student assessments require explicitly stated 'intended outcomes' be achieved. The intention being to allow students to direct their learning efforts towards 'outcome' attainment and to guide assessors on what they are supposed to measure via assessments. The Code provides guidelines for MET providers working under the directives of the National Maritime Regulators to interpret the STCW Code requirements and develop the seafarer training curriculum (with the intended outcomes) to assure that students can demonstrate the attainment of the minimum standards of competence established by the STCW Code.

This paper argues that the STCW Code fails to provide explicit guidelines and instead lays down vague statements which can encourage individual interpretations as to what benchmarks should guide competence assessment. If the benchmark falls short in the measure of essential, minimum, and required competence, graduating seafarers may lack the required competence to perform in a consistent manner in the workplace. This can be dangerous for the shipping industry where any regional weakness in assessment against the STCW Code has profound ramifications as it is an international industry where employees are sourced globally. The perceived oversight of the STCW Code continues into the lack of essential 'criteria and gradations for the quality of performance', and 'context', which can describe the student performance and contextualise the evidence of competence produced. Sub-standard evidence diminishes the value of the resulting CoC creating dissatisfaction among the concerned stakeholders, such as the employers (Cox, 2009; Cross, 2007; Pecota & Buckley, 2009; Sampson, Gekara and Bloor, 2011). Supported by specific examples from the Code, basis will

be laid to highlight the need for a review and improvement to the STCW Code as a standard with unambiguous, assessable outcomes. Additionally, a review of literature in the area of authentic assessment will be used to provide theoretical arguments in support of its use to address the inherent flaws in the STCW Code and improve upon the resulting training outcome.

2. Structure of the STCW Code for 'standards' for competence assessment

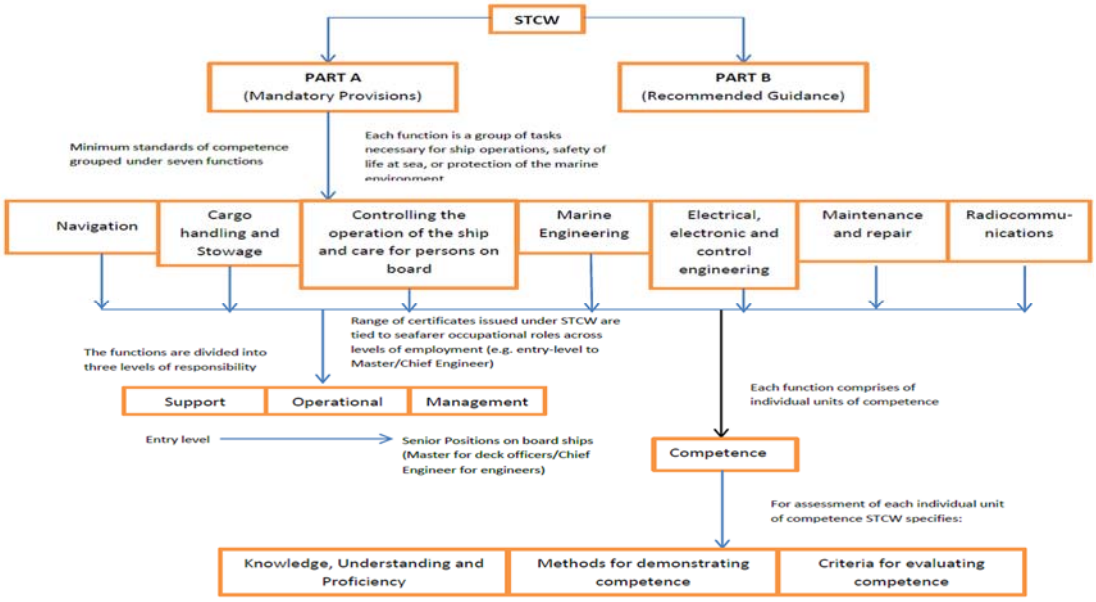


Figure 1: Standards of Competence and Assessment as laid out in the STCW Code

Figure 1 provides a snapshot of how the STCW Code is currently structured in providing 'standards' for competence assessment. As can be seen in the figure, the 'standards' are grouped under seven functions for the three levels of responsibility. Table 1 shows that under these seven functions, the competence for every individual task (or unit of competence), the Code specifies the minimum knowledge, understanding, and proficiency. The evidence of having achieved the required standard of competence is provided with the methods for demonstrating competence and the criteria for evaluating competence.

Competence	Knowledge, understanding and proficiency	Methods for demonstrating competence	Criteria for evaluating competence
Carriage of dangerous goods	International regulations, standards, codes and recommendations on the carriage of dangerous cargoes, including the International Maritime Dangerous Goods (IMDG) Code and the International Maritime Solid Bulk Cargoes (IMSBC) Code Carriage of dangerous, hazardous and harmful cargoes; precautions during loading and unloading and care during the voyage	Examination and assessment of evidence obtained from one or more of the following: .1 approved in-service experience .2 approved simulator training, where appropriate .3 approved specialist training	Planned distribution of cargo is based on reliable information and is in accordance with established guidelines and legislative requirements Information on dangers, hazards and special requirements is recorded in a format suitable for easy reference in the event of an incident

Table 1: Example within the STCW Code stipulating the minimum standards of competence for assessment (Source: STCW, 2011)

3. STCW lacks explicit guidelines for 'outcome' development

Assessment is a significant component of education along with learning and teaching as it provides feedback about a student's progress and achievements, the effectiveness of the teaching and instruction methods, and the course outcomes [University of Tasmania (UTAS), 2011] while supporting the overall goal of improving student learning (Palomba and Banta, 1999). One of the functions of assessment is also to gauge whether the student has achieved the desired outcomes that the learning tasks and teaching processes intended. These outcomes or more correctly 'learning outcomes', define what the students should be able to do at the end of a learning period (Driscoll and Wood, 2007). The learning outcomes thus guide the teaching and instruction towards the assessable outcome. Hence, outcome statements should always precede assessments (Wiggins, 1998). The outcomes should be explicitly stated to ensure that the appropriate assessment methods are adopted to produce the required evidence of outcomes achievement. For example, if the intended outcome is to develop a student's professional competence to fight fires, then the evidence of such competence will be more credible and valid via practical demonstration and not just rely on written examinations. On the reverse side, if the outcome to be achieved was a students' ability to recall the theory behind the cause of fires, written and oral examinations may be more appropriate than practical drills and exercises.

Although, the STCW Code is not curricula or a source of learning outcomes, the 'standards' provided in its Code guide the MET providers (working under the regulatory bodies) to develop curricula with learning outcomes. However, Table 2 provides an example of how at times the 'standard' in the STCW Code can only be a 'standard' of what the students should know in terms of content with some suggested indicators of competence instead of providing 'standards' of demonstrated performance. This makes the STCW an input-based (CEDEFOP, 2010) standard, which is in direct contradiction to the 'outcome-based' objective of the STCW'95.

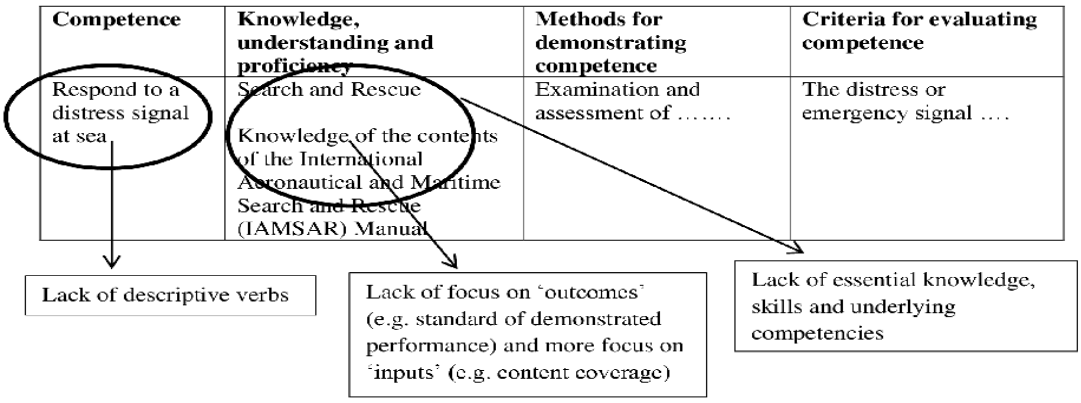


Table 2: Extract from the STCW Code for the function of navigation at the operational level (STCW, 2011)

The word 'respond' represents an action word that provides some indication of what the students should be able to do but does not provide the reader with a 'standard' of how well they must do it. The task should be described with a verb that provides qualitative and/or quantitative descriptions of specific 'performance standards' expected from the students. Descriptive verbs can accurately describe the 'action' outcome expected during student assessments and ensure that the teaching and instruction process for students follow accordingly allowing them to learn and practice the required skills.

The 'standard' for demonstrated performance in the STCW Code should ideally also identify some of the essential knowledge, skills, and behavioural attributes required to perform the task at a professional level. It is because developing the professional competence to perform the task necessitates both cognitive ability to recall information (knowledge) and apply it (skills) based on analytical and critical thinking (Nusche, 2008). Underlying it are the principles, values, and attitudes (behavioural attributes) that are non-cognitive skills developed by the profession through historical experience that promote reflection and shape thought and prompt responses across a range of contexts (Moore and Asay, 2013). However, Table 2 reveals that the STCW Code fails to identify such essential elements.

Lack of descriptive verbs that provide specific and measurable performance standards as well as a lack of essential knowledge, skills and behavioural attributes leave it to the discretion of the National Maritime Regulators and MET providers to develop them. This creates a risk of individual interpretation which in some cases may lead to subversion by low standards and expectations (Wiggins, 1998). If MET providers set low expectations for their students that do not reflect workplace standards, the seafarer may hold a CoC but lack the required level of competence. This can prove dangerous for employers that trust such seafarers with ships worth millions of dollars putting the lives of other seafarers and passengers sailing on these ships and the marine environment at risk.

4. STCW lacks explicit 'Criteria' and 'Standards for Criteria'

Driscoll and Wood (2007) describe criteria as the essential qualities expected from a student's performance that allows them to demonstrate and provide evidence of the achievement of learning outcomes. For example, in Table 3, the 'criteria' column identifies that essential criteria for the 'prevention, control and fighting of fires' is the identification of the type and scale of the emergency.

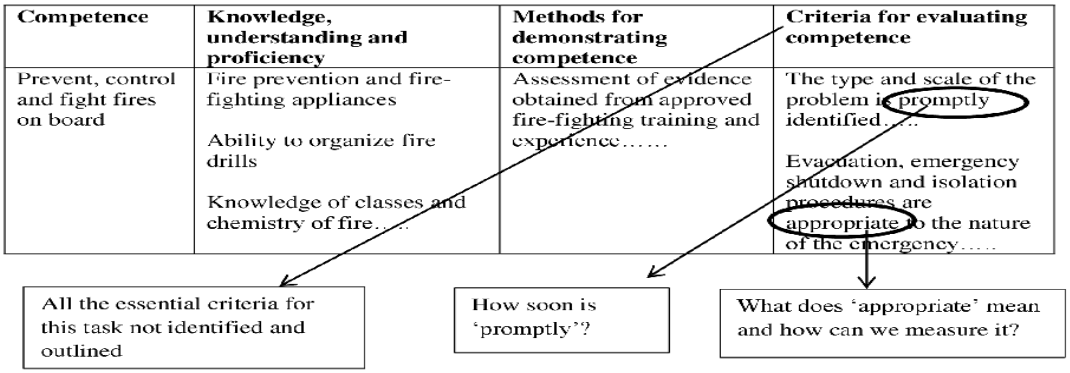


Table 3: Extract from the STCW Code for the function of controlling the operation of the ship and care for persons on board at the operational level (Master and deck department) [24]

However, Table 3 can also be used as an example to show that the STCW Code may overlook some of the other essential and necessary criteria required for performing the tasks listed. For example, the essential and mandatory criteria for donning and operating personal protective equipment for fighting fires are essentially missing from the 'criteria' column. Additionally, the column may also lack in providing a definitional glossary defining key words that shape assessment for the essential criteria. For example, the use of the words 'promptly' and 'appropriate' do not explain how 'quickly' or 'accurately' the task is to be performed. What is the measure that indicates competence as per workplace standards? Lack of measures may lead to vagueness for students and assessors on what is to be expected from the performance. The criteria should describe such words in measurable terms across a range of cultural and performance contexts, e.g. timeframes.

This would not mean that students who are unable to perform the task in the stated timeframe will be deemed incompetent. The criteria should be explained by a range of performance levels that provide a gradation of the quality of performance (Andrade, 2000) or an accurate description of the current competence of the student in performing the task. For example, to perform the task identified as essential criteria in Table 3 ('type and scale of the problem is promptly identified…') the gradation of the quality of performance could be written as,

Criteria	Standard 1 (Deemed insufficient to be declared competent at any level)	Standard 2 (Minimum required to be deemed competent at support level)	Standard 3 (Minimum required to be deemed competent at operational level)	Standard 4 (Minimum required to be deemed competent at management level)
Identify the type and scale of the problem	Type and scale of the problem identified in less than ..... minutes of observation or when made aware of	Type and scale of the problem identified in less than ..... minutes of observation or when made aware of	Type and scale of the problem identified in less than ..... minutes of observation or when made aware of	Type and scale of the problem identified in less than ..... minutes of observation or when made aware of

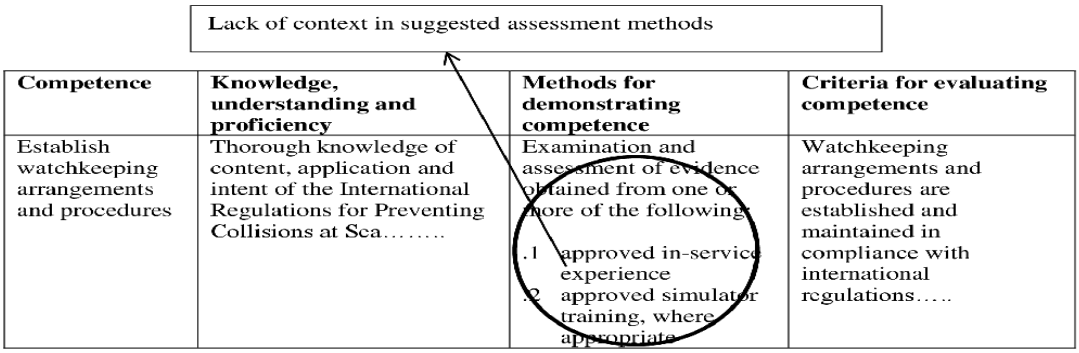
Table 4: Example of how the STCW Code could define the gradations for the quality of student performance

Explicit 'criteria' and 'gradations of the quality of performance' expected from students are essentially missing from most of the tasks described in the STCW Code.

5. STCW lacks explicit 'Contexts'

Forneris and Peden-McAlpine (2006) define context as the foundation upon which a learner's knowledge is constructed in an environment that includes culture, underlying assumptions, previous knowledge, facts, rules and principles. Statement about students' performance made in the specific context in which the assessment was carried out, may inform stakeholders whether competence developed can be directly transferable to workplace or not. For example, the competence to plot a ship's position on a hydrographic chart using GPS data in a classroom may be directly transferable to the workplace (termed as transferable skills), whereas the competence to manoeuvre a vessel which was developed in a simulator may not (termed as non-transferable skills).

Table 5 shows that the STCW Code provides the 'methods for demonstrating competence' but does not explain the contexts in which such demonstration should be carried out. For example, in the case of 'approved in-service experience', should the evidence of competence be collated when the vessel is at sea, at anchor or alongside a port? Should the watchkeeping be done alone or under observation of an onboard assessor? Similarly, the Code does not explain what kind of simulations the simulator should create to obtain reliable evidence of competence. Should the simulated scenarios comprise of other ships to assess the students' competence to apply the relevant theoretical knowledge?



Competence	Knowledge, understanding and proficiency	Methods for demonstrating competence	Criteria for evaluating competence
Establish watchkeeping arrangements and procedures	Thorough knowledge of content, application and intent of the International Regulations for Preventing Collisions at Sea……..	Examination and assessment of evidence obtained from one or more of the following: .1 approved in-service experience .2 approved simulator training, where appropriate	Watchkeeping arrangements and procedures are established and maintained in compliance with international regulations…..

Table 5: Extract from the STCW Code for the function of navigation at the management level [24]

In an ideal world, educational institutes would exactly replicate workplace situations as the skills developed could then be directly transferred to the professional world. However, due to the complex nature of the ship as a workplace, it may be difficult for MET providers to recreate exact workplace settings. In such cases, assessments should be designed to contextually resemble situations likely to be faced by students in the real world, thus making it apparently 'real' rather than apparently artificial (Cumming and Maxwell, 1999). For example, to demonstrate their ability to manoeuvre ships, seafarer students may find simulators more 'real' than a decontextualized environment of a classroom. However, the context in which the assessment is carried out should be clearly defined. Although, the evidence produced from such assessments will not be an accurate reflection of professional competence, it will provide a contextual evidence of competence. Such evidence informs the concerned stakeholders (e.g. employers) about the gaps in the knowledge and skills between those that have been covered and assessed through learning outcomes and transferred to the workplace, and those that can only be truly acquired in the employers' context.

Certainty as to expectations and the actual standard of performance for a graduating student is essential. The STCW Code should explicitly describe the contexts, under which the students' ability to perform the tasks should be assessed. A lack of descriptive contexts may lead to assessors using their individual interpretation in creating contexts for assessments. Different contextual scenarios and varied contextual evidence will complicate matters for the employers and training providers intending to fill the gap. Individual training needs would have to be determined for the employees, which in some cases, may lead to extra training costs and loss of time. Costs of additional training are often borne by the employer (Hanzu-Pazara and Arsenie, 2007). Although employers have training obligations for preparing their employees for specific types of vessels, costs borne for aimless training should be avoided as it can cause a significant impact on the employers' budgets and timelines.

6. Why the STCW needs 'Authentic Assessment'

Although it is imperative that the attention of the IMO is drawn to the inherent flaws of the STCW Code to enable its revision, such processes cannot be expedited. The IMO provides a critical 'safety net' and has a well-established process whereby revisions occur due to the global nature of seafaring. Nevertheless, revisions are required. However, the current focus should be on addressing the issue of the STCW Code failing to provide explicit guidelines to the MET providers and the regulatory bodies for developing learning outcomes that is standardized globally. Interpretation of vague guidelines left to the discretion of the individual parties may cause teaching, learning, and assessment process to vary widely on a global scale leading to inconsistencies in the development of competence level of seafarers. One solution that can be explored to address this problem is the use of authentic assessment methods.

The examples of excerpts from the STCW Code provided in this paper suggest that the Code lacks guidelines for the development of outcomes that can enable students to develop the necessary skills to perform to workplace standards. Authentic assessment tasks are uniquely characterised by tasks contextual to the workplace situations (Herrington and Herrington, 2006) that will replicate the complexities and challenges students will confront in the real world (Frey, Schmitt and Allen, 2012). Performing tasks with strong figurative contexts and fidelity to ship-based scenarios will develop the necessary transferable skills to a higher level of reliability and validity than completely decontextualized scenarios, which may be currently permitted by the STCW Code. Due to the complexity in exactly recreating a ship-based environment, authentic assessment tasks used in seafarer education may only have contextual resemblance to workplace scenarios. Hence, some of the skills developed may not be directly transferable to the real world. However, authentic assessments provide opportunities for students to frequently reflect (Herrington, Reeves and Oliver, 2010) on their work to recognize gaps in their knowledge and grasp cues for enhancing transfer of context free transferable skills and domain specific non-transferable skills (Wiggins, 1993).

Students cannot develop workplace skills by a one-off performance in authentic assessments. To perform to workplace expectations and develop the skills, students should be provided with prior opportunity to practice the skills under guided instruction and teaching. Authentic assessments have been characterised to not only guide the assessment process but also be designed to be a continuous process integral with the learning and teaching, which will allow students to practice skills till they reach the required level of competence (Benner, 2001). Gulikers, Bastiaens, and Kirschner (Gulikers, Bastiaens and Kirschner, 2004a; 2004b) suggested a five-dimensional framework (the assessment task, the physical context, the social context, the assessment result or form, and the assessment criteria) for designing authentic assessment with pertinent questions being framed to consider different dimensions. The framework requires the 'task' that represents professional practices be explicitly defined. It compels assessors to think about the outcome and the required evidence that has to 'result' from or 'form' the basis for the assessments. Such a framework for authentic assessment that requires explicit description of the task to be performed by the student, and the evidence that reflects the level to which it was performed, may provide the contextual evidence that is not currently promoted by the STCW Code.

The framework also requires description of contexts (physical and social) under which the task is to be performed. Explicit descriptions of contexts under which student assessments should take place are essentially missing from the STCW Code. Physical and social context description for authentic tasks should ideally reflect how closely the assessment resembles the professional world (Gulikers et al., 2004a). For example, students should not only have access to resources normally available to them on ships during tasks, but the resources should be applied to reflect the way knowledge, skills, and underlying competencies will be used in the real world. Students' are then assessed on their ability to integrate different competencies that may develop their critical thinking and higher order cognitive skills (Walczak, 1999). Such assessments are a move away from examinations that only require memorized responses to questions. Ability to memorize is a lower level cognition which is not sufficient for performing in workplaces such as ships, where a higher level of cognition is required to assimilate, analyse, and structure (Wiggins, 1990) information for decision making and problem solving. Seafarers who are trained to rely on memory and not to assimilate and analyse the available information to deal with routine or novel situations in the context of the work environment, may suffer from memory failure leading to human errors (Prasad, 2011).

Finally, the framework designed by Gulikers et al. (2004a) focuses on designing the assessment criteria. 'Criteria' in this case refers to the basis on which the evidence of student work produced from the assessments, is judged. Setting the assessment criteria may also guide the learning process as the seafarer students will have a clear understanding of what is expected during the learning process and during their assessments. In authentic assessments, students have access to the performance criterion (reflecting workplace requirements) beforehand (Lund, 1997) for them to aim for the desired level of performance, ensuring that they possess at least the minimum competence level essential for the workplace at a particular level of responsibility. Designing the assessment criteria will require identification and outlining of essential qualities (or underlying competencies) expected from a student during the task performance. Additionally, it will also require describing levels that can define the different gradations of quality of performance. Such requirements are currently lacking and not promoted by the STCW Code.

7. Conclusion and the Way Forward

The STCW Convention led to the development of the STCW Code to provide global, minimum standards of competence for seafarers. The 'standards' were expected to act as guidelines for regulatory bodies and MET providers worldwide to develop consistent and uniform training outcomes. However, the paper argues that the STCW Code is too vague and this may lead to individual interpretation in adopting learning and assessment processes towards competence development, which creates the risk of seafarers graduating with CoCs but lacking the required competence for workplace operations. The seafaring industry sources its employees globally and cannot afford to operate under such risks. Based on the review of the selected excerpts, the STCW Code gives the impression of being an input-based education system and not an outcome-based as it was originally intended to be. An input-based system may prove to be regressive for the seafaring industry due to its focus on curriculum and content

coverage and not on the appropriateness, learning and assessment, or the attainment of the desired competence outcomes by the student. Graduates may be assessed as competent but lack the necessary attributes making it a point of risk for employers and governments relying on MET providers to deliver seafarers that meet the required standard.

Based on a brief review of some of the past literature on authentic assessment, this paper suggests it as one of the possible solutions to address the discussed weaknesses of the STCW Code. Although the review is not comprehensive of all the literature, the paper discusses the ideas of major authors on authentic assessments, such as Wiggins (1993), Gulikers, Bastiaens and Kirschner (2004a; 2004b) and Herrington, Reeves and Oliver (2010). Based on their ideas, it is suggested that assessment tasks that contextually resemble real world situations may not only engage students in learning but also assist in the development of skills which may be directly transferred from MET environment to workplace settings. For non-transferable skills, it is suggested that authentic assessment may allow assessors to contextualise the competencies of the seafarers. This would allow stakeholders to identify the gaps, if any, between competence developed in educational settings and those required at the workplace, to be filled with additional training. In the absence of such contextual evidence, any additional training provided to employees is a 'risk accepting' behaviour that is more about 'hope' than assurance that a standard of performance has been obtained. However, further investigations requiring collection of empirical data is needed to substantiate theoretical claims stating that authentic assessment may improve the STCW training outcomes and the resulting training mandate.

7.3. Paper III

USING AUTHENTIC ASSESSMENT TO ENHANCE SEAFARER STUDENT ENGAGEMENT AND THEIR ABILITY TO TRANSFER LEARNING

Samrat Ghosh
Australian Maritime College, University of Tasmania
Marcus Bowles
Australian Maritime College, University of Tasmania
Dev Ranmuthugala
Australian Maritime College, University of Tasmania
Benjamin Brooks
Australian Maritime College, University of Tasmania

Abstract

Past research shows that seafarer employers are critical of some of the assessment methods adopted by the educators at maritime education and training institutes (METs) to assess the competence of seafarers. The criticisms included the failure to develop and assess the holistic skills required to deploy competence in a range of contexts. Moreover, the decontextualized scenarios used in assessment methods disengage students from the learning process as they fail to recognize the significance of learning to the real-world. This paper argues that seafarer students can be engaged through authentic assessments conducted in real-world contexts that will test their ability to put theoretical knowledge developed in classrooms to practical settings resembling workplace scenarios. The arguments are based on the theories of constructivism and self-efficacy that underpin the concept of authentic assessment. The theories are used to explain greater student engagement through involvement in the process of knowledge construction that also develops metacognitive skills for the transfer of learning to different contexts. The theoretical arguments are supported with empirical evidence from past research to provide a robust justification for the use of authentic assessment in seafarer training to obtain similar outcomes.

Keywords – Seafarer, authentic assessment, student engagement, learning transfer

1. Authentic assessment promotes student engagement

Student interviews during the course review of deck officers (Australian Maritime College (AMC), 2011) carried out in the Australian Maritime College (AMC) revealed that seafarer students felt disengaged with traditional assessments when used for assessing their competence to perform workplace tasks. Traditional assessment methods such as pen and paper testing, oral exams, and multiple choice questions (MCQs) may be effective in assessing lower order cognition skills of memorising and ability to regurgitate which is necessary but not sufficient for performing in workplaces such as ships, where a higher level of cognition is required to assimilate, analyse and structure (Wiggins, 1990) information for decision making and problem solving. For example, Table 1 shows how seafarer assessments may use a combination of traditional and authentic assessments to assess competence to perform STCW tasks to workplace standards.

Competence	**Knowledge, understanding and proficiency**	**Methods for demonstrating competence**	**Criteria for evaluating competence**
Respond to a distress signal at sea	Search and Rescue Knowledge of the contents of the International Aeronautical and Maritime Search and Rescue (IAMSAR) Manual	Examination and assessment of evidence obtained from practical instruction or approved simulator training, where appropriate	The distress or emergency signal is immediately recognized Contingency plans and instructions in standing orders are implemented and complied with

Ability to respond to distress signals to workplace standards should be assessed via authentic assessments

Knowledge-based components of competence may be assessed via traditional pen and paper, MCQs, or oral examinations

Table 1: Extract from the STCW Code'95 for the function of navigation at the operational level (STCW, 2011)

However, an ethnographic case study involving a small sample of 16 students carried out by Emad and Roth (2007) in a Canadian maritime institute, revealed that students were aware of the fact that traditional exams comprised mainly of the questions which were drawn from a question bank. Over time they could predict the range of questions and prepare accordingly. Such assessments that lack innovation in design can encourage memorizing to pass examinations instead of the deployment of critical thinking and problem-solving skills that are essentially required at every workplace.

Assessments designed to assess professional competence of seafarers to perform real-world tasks should ideally create similar scenarios for student performance. For example, student interviews at the AMC (AMC, 2011) revealed their preference for assessments that are contextually similar to challenges found at workplace, in order to relate classroom learning to professional practice. Lack of contextual similarity in learning and assessment makes it difficult for students to relate how

skills and knowledge developed in classrooms can be applied in workplace contexts (Findlay, 2013). It also raises the questions of validity where students start to question the relevance of the assessments and the competence it purports to assess, thus disengaging students. Assessments that are designed in real-world contexts are defined as authentic assessments (Wiggins, 1990; Archbald, 1991; Gulikers, Bastiaens, and Kirschner, 2004; Darling-Hammond and Snyder, 2000).

Meaningful contexts through real-world scenarios create high level of student engagement and commitment. For example, interview of six students in a study by Richards Perry (2011) revealed students' preference for meaningful and relevant learning experiences; and authentic learning. The convenience sample was kept to a low number of six students to gain a greater depth of inquiry through an extensive interview protocol. However, to be engaged in learning, students will not only require meaningful contexts but also to be active participants in the knowledge construction process that precedes the assessments (Hart et al., 2011). Although the uniqueness of authentic assessment lies in the setting of tasks in real-world contexts, drawing upon the literature (Wiggins, 1990; Archbald, 1991; Darling-Hammond and Snyder, 2000; Gulikers et al., 2004; Gulikers, 2006), authentic assessment herein will encompass:

tasks *resulting in outcomes* in a *real world context* that require an *integration of competence* to solve *forward looking questions* and *ill-structured problems*; **processes** that require *performance criteria to be provided beforehand* and *evidence of competence to be collected by the student*; and **outcomes** that result in *valid and reliable student performance, contextual and multiple evidence of competence, higher student engagement, and transfer of skills to different contexts.*

Traditionally, seafarer education has been teacher-centric where students have been passive receivers of knowledge (Lewarn, 2002). This does not allow seafarer students to become active participants in the learning process. In active learning, students are not only mere receivers of knowledge but also involved in the construction of it. According to the learning theory of constructivism, construction of knowledge allows students to develop a deeper understanding of the learning content. Authentic pedagogical practices are influenced by the constructivist philosophy of student-centred learning where students create meaningful knowledge in real-world tasks (Morrissey, 2014), thus engaging students in the learning process. For example, a study by Quartuch (2011) showed that the use of authentic assessments allows students to become civically engaged demonstrating key content knowledge, critical thinking, and understanding complex issues from multiple perspectives. However, these findings are reliant on a small sample of 11 students from a 12th grade college preparatory American Government and Economics class in a large urban high school in eastern Pennsylvania.

Seafarer students are expected to achieve learning outcomes driven by the Standards of Training, Certification, and Watchkeeping (STCW) Code. However, lack of descriptive outcomes in the Code (Ghosh et al., 2014a) and traditional teaching practices often do not provide the students with clear expectations of learning standards to be achieved. In authentic assessment, the teacher provides a roadmap of the entire subject to be learned while allowing students to construct their understanding of the topic. Having standards of performance provided beforehand, would provide opportunities to seafarer students to reflect on their learning and carry out self-assessments of their thinking and practices towards achievement of the required standards.

As learning progresses, learners assume increasingly more control over the sequence in which they want to engage their learning (Schell, 2000) and gain mastery over knowledge and skills learnt through strategic and critical thinking (Fredricks and McColskey, 2012). For example, a study by Findlay (2013) revealed that relationship between student-teacher based on the qualities of authenticity, belief, empowerment, and life-long learning, enhances student motivation and engagement. While the student and teacher relationship in authentic teaching was found to create a positive learning environment; belief, empowerment, and life-long learning was promoted through student reflection and self-assessment achieved through self-efficacy in the constructivist view of learning.

'Meaningful reflection' allows individuals to reflect on acquired knowledge in different situations encouraging them to become life-long learners (Schon, 1983). The technical term for this type of reflective process is metacognition (Scott, 2000). Metacognitive reflection and self-assessment teaches students to identify the gaps between their current competence and those required by educators or employers at the workplace (Boud and Walker, 1998).

This is a key requirement for transfer of learning to take place (McCarthy, 2013).

2. Authentic assessment promotes transfer of learning

Official investigations and analysis of marine accidents have revealed that seafarers assessed as competent in the use of particular skills in a context have failed to apply them in another (Pecota and Buckley, 2009). Although reliant on a small sample, a study that comprised of a series of seventeen interviews with employers in the UK, Philippines, and Singapore, Sampson et al. (2011) discovered that employers were critical of some of the current assessment methods in use for seafarer assessment. The interviewees were fleet personnel managers from both owner operator companies and of ship management companies, involved in the employment of seafarers largely from countries like India, Myanmar, and Philippines and from Eastern Europe. According to the employers, current assessments assess a limited range of job specific skills (Cox, 2009; Cross, 2007), in settings that provide insufficient cues to the students on how the competence acquired in classrooms can be used in different contexts.

Students who are able to frequently reflect on their learning to recognize gaps in their own construction of knowledge and improve on them, begin to grasp cues on applying the same knowledge (developed in a specific context) to different contexts (Leberman, 1999) causing a transfer of learning (Donovan, Bransford, and Pellegrino, 1999) Authentic assessment are formative assessments that provide students with frequent opportunities to reflect (Herrington, Reeves, and Oliver, 2010), acting as a "pit stop" where students and assessors can reflect on the application of their skills (Curry, Caplan, and Knuppel, 1994) in a particular context and identify additional training requirements for different contexts.

Metacognitive reflection and self-assessment during construction of knowledge have been shown to increase the degree to which students will transfer to new situations without the need for explicit prompting (Bransford, Brown, and Cocking, 2000) For example, a study by Sator (2000) showed that metacognitive reflection as a thinking skill was evoked by all the reflection exercises in the

Skills Transfer learning module of a bridging online course. The course was part of an online co-operative education (learning strategy that provides a structured method for bridging academic learning with practical experiences in the workplace) preparatory curriculum where the 28 participating students revealed strong evidence of metacognitive reflection in strategies adopted for successful transfer. The study involved a qualitative content-analysis of online discussion to understand if the thinking skills exhibited were consistent with the understanding of bridging techniques that support transfer of learning.

Seafaring assessments are usually summative carried out at the end of the learning period, not allowing the students to engage in deep reflection during the assessment process. Implementing formative authentic assessments would allow seafarer students to engage in metacognitive reflection to recognize the gaps that exist in their understanding. As gaps are recognized and become significant to students, they may locate, apply, and connect previous learning as well as new knowledge (Scott, 2000) and skills causing transfer of learning.

According to the self-efficacy theory by Bandura (1977), construction of knowledge as promoted by authentic assessment, develops critical thinking skills enabling students to re-evaluate their learning, causing behavioural changes that promote positive growth in cognitive development which can be used to assimilate, analyse, and structure information for decision making and problem solving as required on ships. For example, through survey of 2567 participants in the graduate studies in education program, Saunders et al. (2001) found positive correlations between authentic assessment and adult learner's cognitive skills. Cognitive development through self-efficacy provides students with the belief and confidence to transfer newly acquired knowledge and skills (Merriam and Leahy, 2005). Learners draw on and extend previously learned knowledge and develop their own cognitive maps to interconnect facts, concepts and principles. As learning progresses, understanding becomes integrated and structured leading students to gain mastery over content (Scott, 2000). Past research suggests that the students' ability to transfer is enhanced when they are able to use the deep understanding of the learning content to interconnect facts and apply it to different contexts (Mestre, 2002).

However, according to the constructivism theory of learning, transfer can be enhanced when learning is contextualized in authentic tasks designed in meaningful contexts (Ertmer and Newby, 1993). Decontextualized learning does not allow students to recognize the connectedness of learning and application of skills developed to the real world which may have a negative impact on transfer (Mbawo, 1995). Due to complexity in recreating ships as workplace on land-based maritime educations and training institutes (METs), most of the learning and assessment in seafarer education takes place in decontextualized scenarios. Transfer is more likely to occur when instructional and application settings are nearly identical (Schell, 2000).

Authentic assessments conducted in real world contexts will provide 'cues' to students on strategies to adopt when performing in the real world. For example, in a study by Herrington and Herrington (1998), six pre-service teachers were assessed at the workplace to study their ability to transfer skills and knowledge developed through authentic pedagogical practices in classrooms. Interviews revealed that all six students had successfully used strategies without the need of prompting from the supervising teacher and attributed their use to the authentic teaching and assessment. Although the findings were derived from a small sample, the emphasis on meaningful authentic contexts in

learning being necessary in preparing students for professional practices was highlighted. Contextualised authentic tasks may not recreate all the conditions of ships as a workplace but may replicate the complexities and challenges faced by seafarers in the real-world.

3. The Way Forward

Students engaged in analysing their own work against pre-established standards of achievement can provide critical feedback to teachers on how engaged the students are (Munns and Woodward, 2006). Concurrently, provision of clear expectations on standards of performance beforehand also allows educators to adopt appropriate pedagogical practices to guide students towards achievement of the desired outcomes (Archbald, 1991), which may improve learning practices. Improving learning practices may allow educators to meet student needs for higher engagement. Students engage in different ways and at times the expectations are not met due to a narrow vision of engagement held by educators (Trygstad, 2010). What may be authentic for educators may not be authentic for students. Educators may require additional training to develop their ability to create authentic assessments that reflect contemporary workplace needs.

In creating authentic ship-based or contextually similar scenarios to assess seafarer competence, educators must investigate the current needs of the employers and workplace expectations.

Student engagement lies more with pedagogical strategies. Contemporary pedagogy that treats students as stakeholders in the educational process may enhance student engagement. Education system must consult those that are designed to serve (Ozimek, 2000). This may allow educators to improve student perceptions of authenticity and in the process enhance engagement that may result from it. Students should be consulted in developing rubrics that provide clear outcomes intended to achieve from the learning process. Students will not only feel involved as an integral part of the learning process but will also provide valuable feedback that may improve the pedagogical practices and the outcomes resulting from it. Educators may also require additional training to develop their ability to create rubrics which can be used for instruction as well as assessment.

Authentic assessment research so far has not investigated its impact on seafarer training outcomes (Ghosh et al., 2014b). Future research needs to investigate if authentic assessments reflecting ships as workplace or contextually similar scenarios can be recreated in METs; and if such practices enhance the ability of seafarer students to transfer their learning from classrooms and simulators to workplace contexts.

Most of the studies used in this paper reveal findings based on a small sample of research participants. Future research needs to corroborate these findings using a larger sample of participants.

4. Conclusion

The literature examined in this paper indicates that students are engaged when they are able to develop a deep understanding and mastery of the learning content. This occurs through meaningful reflection and self-assessment; actively involved in the construction of knowledge; and form a deep connectedness of learning with real-world applications. In this way, student engagement can be used as an indicator to measure the quality of learning and teaching in universities.

Traditional learning and assessment methods in seafarer education are largely failing to engage the students in learning; and developing their ability to transfer learning acquired in classrooms to workplace settings. This paper presents a shift from teacher-centric education as currently practiced in seafarer education to learner-centric authentic pedagogical practices, as a possible solution. Student-centred education allows the students to be part of the knowledge construction process where they are fully aware of the learning expectations from the beginning of the learning period. This reduces uncertainty with assessment practices and creates higher student engagement. High student engagement promotes deep understanding of learning content and motivation to master skills and knowledge. Students develop a higher cognition to relate previous learning and newly acquired knowledge to apply in different contextual settings, which may assist in transfer of competence. The ability of students to perform workplace tasks is then not limited to specific classroom or simulator scenarios, as found currently with seafarer students.

However, contexts of learning and assessment needs to be meaningful for students to relate classroom learning to real-world practices as decontextualized scenarios make transfer nearly impossible. Such meaningful contexts can be achieved through real-world scenarios in authentic assessment. Future research needs to establish if authentic assessment in seafarer training can enhance student engagement and their ability to transfer learning to different contexts, making graduates more competitive in a global shipping world. As most of the studies used in this paper reveal findings based on a small sample of research participants, future research on authentic assessment in the area of seafarer training needs to consider larger number of respondents.

6.1. Paper IV

AUTHENTIC ASSESSMENT IN SEAFARER EDUCATION: USING LITERATURE REVIEW TO INVESTIGATE ITS VALIDITY AND RELIABILITY THROUGH RUBRICS

Samrat Ghosh
Australian Maritime College, University of Tasmania
Marcus Bowles
Australian Maritime College, University of Tasmania
Dev Ranmuthugala
Australian Maritime College, University of Tasmania
Benjamin Brooks
Australian Maritime College, University of Tasmania

Abstract

With the Standards of Training, Certification and Watchkeeping Convention 1995 (STCW'95) moving seafarer training towards outcomes-based education (OBE), emphasis has shifted to assessment practices that will allow seafarer students to demonstrate their ability to perform workplace tasks at standards described in the STCW Code. This paper argues that authentic assessment comprising of performance-based tasks applied in real-world and meaningful contexts, can provide a holistic approach to competence assessment for seafarers. But authentic assessment can capture essential aspects of workplace tasks and result in consistency of student performance in different contexts only if they are valid and reliable. Rubrics as assessment tools are known to increase validity and reliability of assessments, however, it can do so only if different aspects of its own validity and reliability have been addressed. A literature review undertaken for this paper has uncovered an absence of academic investigation and empirical study on the different aspects of validity and reliability of authentic assessment through assessment rubrics. Moreover, there exists an even greater absence of global research on authentic assessment in the area of seafarer training. Through an investigation of authentic assessment, this research has uncovered the importance of using valid and reliable rubrics in order to improve not only the assessment process but also the tools and methods used to support the valid, reliable, and authentic assessment of outcomes achieved in the learning process. Future research aims to offer insights into improving the validity and reliability of rubrics and to empirically investigate how they can be used in authentic assessment within the confines of the STCW Code, in particular to improve: seafarer training practices, student engagement, resulting learning outcomes, and employer and regulator satisfaction with the attainment of the standards stipulated in the STCW Code to produce an evidence of competence.

Keywords – Authentic assessment, seafarer education and training, rubrics, validity, reliability

1. Introduction

In education, assessment can be defined as "a systematic collection, review, and use of information" (Walvoord, 2004) to acquire feedback about: a student's progress and achievements, the effectiveness of teaching and instruction, and the attainment of course outcomes (University of Tasmania (UTAS), 2011), while fulfilling the overall goal of improving student learning (Palomba and Banta, 1999). In Outcomes-Based Education (OBE) such as Vocational Education and Training (VET) or Competency-Based Training (CBT), assessments also provide feedback about the attainment of minimum standards by students that are essentially required for the workplace (Brady, 1997; p.10). Standards in such cases become the outcomes (Burke, 2011) or more correctly 'learning outcomes' establishing what the students should be able to demonstrate at the end of the learning period (Driscoll and Wood, 2007). Students direct their learning efforts towards 'outcome' attainment and assessors are guided on what they are supposed to measure via assessments. The evidence produced from the assessments can be used by educators to not only improve teaching practices by identifying learning needs, but also to meet accountability requirements by providing feedback to stakeholders on the learners' progress towards achievement of standards (Brindley, 1998).

Standards for the occupational practice of seafaring are provided through the Standards of Training, Certification and Watchkeeping (STCW) Code of the STCW Convention that was introduced by the International Maritime Organization (IMO) in 1978 (then known as STCW'78). The STCW'78 was essentially knowledge-based comprising a syllabus for a quantifying examination instead of focusing on skills and abilities necessary to perform workplace tasks (Morrison, 1997). The IMO revised the STCW Code through the 1995 amendments (since known as STCW'95) intending to fundamentally improve the training mandate by making it outcome-based. As a requirement of OBE and for the purposes of the certification and licensing, seafarers are required to demonstrate the achievement of the STCW standards through assessments.

Demonstration of attainment of competence that resembles workplace standards may require assessments that not only assess students' progress against outcomes attainment but also their ability to perform workplace tasks. Evidence produced through traditional assessment tasks such as multiple-choice questioning or oral examinations can provide indicators for students' mastery of content knowledge but may not be able to adequately capture different aspects of a complex student performance resembling workplace tasks (Montgomery, 2002). Such performance can be captured through assessment rubrics which comprise of individual and essential dimensions of performance known as criteria along with standards for levels of performance against those criteria (Jonsson and Svingby, 2007). Rubrics involve creating a standard and a descriptive statement that illustrates how the standard is to be achieved (Cooper and Gargan, 2009). Rubrics may report on outcomes attainment, but the validation of attainment is achieved through the assessment process (Davis et al., 2007).

To determine if the intended outcomes have been achieved and to collate evidence of the same, assessors need to decide whether the selected assessment methods will adequately allow for evaluation and demonstration of the students' learning outcomes (Moskal, 2000). The quality

of the information provided on outcomes attainment by the rubrics will only be as good as the assessments on which the reporting is based (Brindley, 1998). The ability to perform workplace tasks should be assessed through assessment methods that resemble professional scenarios. Hence, fidelity of context to conditions in which the professional skill would be applied becomes an important element of assessment methods adopted. Such performance-based assessments applied in real-world contexts have often been described as authentic assessments (Herrington and Herrington, 1998; Reeves and Okey, 1996; Wiggins, 1993; Meyer, 1992).

However, fidelity of context cannot alone assure that essential aspects and constructs of professional competencies are being accurately assessed. Assessments should be valid and reliable to do so. Validity refers to the extent to which the evidence produced through assessments supports the inferences made about the student's competencies and whether such inferences are being interpreted in appropriate contexts (Moskal and Leydens, 2000). On the other hand, reliability refers to the consistency of assessment scores obtained every time the same competencies are assessed irrespective of the scorer, time period between the assessments, and the contextual and individual learning variables under which the assessments occur (Moskal and Leydens, 2000). Rubrics provide clear statements on learning and performance expectations for both educators and students. Such statements can then be used to assess if intended outcomes were achieved by students, educators, and assessors. Hence, rubrics are highly regarded as tools that increase validity and reliability in assessments (Rezaei and Lovorn, 2010; Jonsson and Svingby, 2007; Silvestri and Oescher, 2006).

This paper establishes the importance of using rubrics as an authentic assessment instrument for assessing outcomes that represent workplace tasks. Authentic assessment is defined collating all the characteristics used by major authors in the field. Validity and reliability are then established as essential criteria for measuring the effectiveness of assessment methods by researchers. Based on an extensive literature review in the area of authentic assessment, this paper explores the practices adopted in the past to improve the validity and reliability of authentic assessment when rubrics are used as an assessment instrument. The review uncovers a lack of holistic approach in addressing both validity and reliability aspects of authentic assessment and an absence of global research on authentic assessment in the field of seafarer education and training.

2. Definitions

2.1 Authentic Assessment

The idea of 'authenticity' in education was conceived and developed in response to increasing accountability to stakeholders. The movement started in the 1980s in the high schools of USA. The term 'authentic' was first linked to student achievement by Archbald and Newmann (1988) requiring them to demonstrate outcomes beyond the school learning environment in an applied/work context. Wiggins (1989) related the term to student assessment while promoting authentic assessment as a process that required student performances (Wiggins, 1990) at standards expected in the professional field. Unlike traditional tests that produced transcripts

with unclear information of actual competence, evidence of student performance at workplace standards would improve accountability to stakeholders.

Authentic assessment is often used interchangeably with performance assessment as it imbibes some of the characteristics of the latter, but they are not synonymous (Marzano, Pickering, & McTighe, 1993). For example, all authentic assessments require a performance of some kind but not all performance-based assessments are conducted in authentic or real-world contexts (Meyer, 1992). Palm (2008) provides a detailed classification of meanings describing the similarities and wide range of differences between the meanings of each concept. Authentic and performance assessments are known as types of 'alternative assessments' to traditional assessments (Dikli, 2003). Traditional assessments include pen and paper testing, multiple choice questioning (MCQs), and oral examinations. Cumming and Maxwell (1999) show that characteristics of authentic assessment can also be found in other assessments, such as problem-based and competency-based assessments but provide clear distinction between them. For example, they explain that authentic assessment is based on theories of learning where performance of tasks occurs in genuine workplace or contextually similar situations. On the other hand, competency-based assessments are based on the theory of vocational education where assessment tasks should represent workplace tasks but can be performed in individual components and not necessarily integrated into one holistic task. Authentic assessments have also been called dynamic assessments (Chance, 1997; Butler, 1999) due to its dynamic nature of evolving to address student learning needs.

This paper defines authentic assessment by collating the characteristics provided by most commonly cited authors in the area (Table 1). The exact number of citations for the individual papers has been obtained from the website of Google Scholar.

Major Author/Year	No. of citations	Context of Study	Real world context	Integration of competence	Known performance criteria	Valid and reliable performance	Forward looking questions	Ill-structured problems	Evidence of competence collected by student	Tasks resulting in outcomes	Contextual evidence of competence	Multiple indicators of competence	Promote student engagement	Allow transfer to different contexts
(Wiggins, 1989)	939	High Schools	✓	✓	✓	✓	✓	✓	✓			✓		
(Wiggins, 1990)	383		✓	✓	✓	✓	✓	✓	✓			✓		
(Wiggins, 1993)	364		✓	✓	✓	✓	✓	✓	✓			✓		
(Wiggins, 1998)	956		✓	✓	✓	✓	✓	✓	✓			✓		
(Archbald, 1991)	27	Schools			✓					✓			✓	✓
(Darling-Hammond and Snyder, 2000)	397	Teacher Education	✓	✓	✓							✓		
Gulikers, Bastiaens, & Kirschner, 2004a	212	Nursing Education	✓	✓	✓					✓	✓			
Gulikers, Bastiaens, & Kirschner, 2004b	17		✓	✓	✓					✓	✓			
Gulikers, Bastiaens, & Kirschner, 2006	33		✓	✓	✓					✓	✓			
Gulikers, Bastiaens, Kirschner, & Kester 2008	23		✓	✓	✓					✓	✓			
(Gulikers, 2006)	10	Nursing Education	✓	✓	✓					✓	✓			

Table 1: Characteristics of authentic assessment defined by most commonly cited authors

Based on the characteristics provided in Table 1, authentic assessment herein will encompass:

tasks *resulting in outcomes* in a *real-world context* that require an *integration of competence* to solve *forward looking questions* and *ill-structured problems*; **processes** that require *performance criteria to be provided beforehand* and *evidence of competence to be collected by the student*; and **outcomes** that result in *valid and reliable student performance, contextual and multiple evidence of competence, higher student engagement, and transfer of skills to different contexts*.

2.2 Rubrics

Rubrics (an example shown in Table 2) are assessment tools that comprise of individual and essential dimensions of performance known as criteria along with standards for levels of performance against those criteria (Jonsson and Svingby, 2007). Although the terms 'criteria' and 'standard' is sometimes used interchangeably, they have distinct meanings (Sadler, 2005). The definitions provided by Sadler (2005) and Spady (1994) provide a robust basis for distinguishing the terms. Standards are defined as levels of definite attainment and sets of qualities established by authority, custom, or consensus by which student performance is judged, whereas criteria are essential attributes or rules used for judging the completeness and quality of standards. Table 2 provides an example of how a rubric may be designed for the unit of competence of "Prevent, control, and fight fires on board" at the operational level from the STCW'95 Code. The move of seafarer training to OBE has shifted the emphasis to demonstration of competence requiring the intended learning outcomes (ILOs) be established and communicated to students beforehand to make the learning process transparent (Biggs and Tang, 2007). As assessment rubrics communicate standards and the feedback for its achievement, they are an essential tool to OBE (Reddy, 2007).

Criteria	Standard 1 (Performance deemed insufficient to be competent at operational level)	Standard 2 (Performance meeting minimum required to be deemed competent at operational level)	Standard 3 (Performance beyond minimum required to be deemed competent at operational level)
Identify the class of fire and choose the correct extinguishing system	Unable to identify the class of fire and/or choose the correct extinguishing system	Able to identify the class of fire and choose the correct extinguishing system	Able to identify the class of fire and choose the correct extinguishing system; Able to identify alternate extinguishing systems for the class of fire
Operate the fire extinguisher as per the manufacturer's instructions	Unable to operate the extinguisher as per the manufacturer's instructions	Able to operate the fire extinguisher as per the manufacturer's instructions	Able to operate the fire extinguisher as per the manufacturer's instructions; Able to demonstrate adoption of measures to prevent the spread of fire and its' reoccurrence
Wear the fireman's outfit as per the manufacturer's instructions and extinguish the fire	Unable to wear the fireman's outfit as per the manufacturer's instructions and extinguish the fire	Able to wear the fireman's outfit as per the manufacturer's instructions and extinguish the fire	Able to wear and use the fireman's outfit as per the manufacturer's instructions and extinguish the fire; Able to demonstrate adoption of measures for the care and maintenance of the fireman's outfit for reuse

Table 2: Example of how a rubric may be constructed for the STCW unit of competence of 'Prevent, control, and fight fires on board' at the operational level

Without rubrics, students have no guidelines towards achievement or to understand the teacher's feedback comments (Montgomery, 2002) on outcomes achieved. For example, using a focus group discussion involving fourteen undergraduate students, Andrade and Du (2005) found the use of rubrics to be very effective in providing performance expectations and feedback about achievement of standards in teacher education. However, using rubrics to communicate standards achieved by students in professional education also requires assessment methods such as authentic assessment that can capture such standards.

Traditional assessments such as multiple-choice questioning and oral examinations assess the ability to recall facts and some of the applied skills (Archbald, 1991) but fail to assess essential behaviour based attributes (Wiggins, 1992) an individual must develop along with technical skills and knowledge that together define professional competence (Sampson and Fytros, 2008). Assessment of professional competence can be captured through authentic assessment tasks based in meaningful contexts and applied in real world or contextually resembling real-

world settings. However, professional competence is developed and assessed under specific contexts in educational settings. Transfer of performance or competence to perform individual components of a task to a holistic performance of the task where integration of competence is required, cannot be assumed (Cumming and Maxwell, 1999). According to Cumming and Maxwell (1999), learning and assessment needs to be contextualised to make it relevant and meaningful for students. Meaningful context can not only provide motivational benefits to student learning but also a clear understanding of learning that can or cannot be transferred to different contextual scenarios. If real life contexts and complexities (task centred approach), cannot be created in assessments, they should then focus on the selected constructs (construct centred approach) of knowledge and skills (Messick, 1996). For example, assessments designed in maritime education and training (MET) institutes may not be able to assess a student's competence to manage large crowds as is required on passenger ships but they may be designed to assess a student's competence to do so through their ability to analyse risks associated with such management or developing crowd management plans. Although such assessments may take place in controlled situations, the authenticity will be reflected through ways in which the same skills would be applied in real-life contexts (Messick, 1994). The standard of learning achieved in the real-world contexts may be communicated via rubrics making it an important authentic assessment instrument for assessing outcomes that represent workplace tasks.

3. Authentic assessment

3.1. Aligning assessment with rubrics

One of the key characteristics requires authentic assessment to provide performance criteria to students beforehand, which can be done through the use of rubrics. Provision of clear expectations of standards of performance via rubrics allows students to learn and educators to adopt appropriate instructional strategies to guide students towards the achievement of the desired outcomes (Archbald, 1991). The use of summative examinations at the end of the learning period represents the final judgement of the students' performance and is often too late to make any changes to the learning strategies. Authentic assessment methods that are based on ongoing use of formative assessments may be more suitable to provide diagnostic feedback and make adjustments to improve the learning process (Burke, 2011).

Hence, the alignment of the learning, teaching, and instruction process towards the achievement of outcomes creates constructive alignment (Biggs and Tang, 2007). Constructive alignment comes from the constructivist theory (Biggs and Tang, 2011), where the student is not a mere receiver of knowledge but is also actively involved in the construction of it while progressing in learning. Newmann et al. (1996); 1995 and Cooperstein and Kocevar-Weidinger (2004) connected authentic assessment to the constructivist way of learning. Although principles of constructivism can allow everyone to construct meaningful learning, Newmann et al. (1996) recommended that high intellectual standards provided through rubrics in authentic assessment can promote highly intellectual construction of knowledge and meaning leading to superior learning and performance that would require students to use higher-order cognitive skills. In the current educational environment of the 21st century, assessments should not only capture the content knowledge or the professional skills

but also higher-order skills (Burke, 2011) of problem-solving, critical thinking, leadership, and team-working. According to Wiggins (1989), assessments should not only monitor standards but also set them to reveal achievement of higher-order skills which may not be quantified but is a necessity in a work context. Traditional assessments are not always performance-based; nor can they be always creatively designed to encourage demonstration of higher-order skills. For example, a study by Brawley (2009) that involved authentic assessment of 24 students in early childhood showed that authentic assessments, when designed properly, are a better way to determine the higher-order thinking skills (as defined by Bloom's taxonomy) required to complete a task. Creating authentic experience for students correctly becomes central to designing authentic assessment.

3.2. Validity and reliability of authentic assessment

Advances in technology such as simulators, web-learning, multimedia, etc. have allowed many researchers (Neely and Tucker, 2012; Neo, Neo, Tan, 2012; Osborne, Dunne, Farrand, 2013; Scholtz, 2007) to use such technology in the area of authentic assessment to create authentic experiences that can replicate real-world tasks for the students. However, Messick (1996) was not convinced that authentic assessments can ever fully represent real-world tasks in educational settings. Messick believed assessments are prone to threats of validity which emphasises the appropriateness of assessment tasks as effective measures of intended learning outcomes (Rhodes and Finley, 2013). Because authentic assessments have a high fidelity to real-world contexts, does not necessarily lead to the conclusion that they are more valid than traditional examinations. Assessment methods should be judged by established criteria for judging the technical adequacy of measures. Key among these criteria are the concepts of validity and reliability (Linn et al., 1991).

Validity and reliability are crucial to the acceptance of authentic assessment (or rubrics as an assessment tool) as an accurate measure of knowledge, skills, and behaviours (Stevens, 2013). There are numerous extraneous variables that affect the validity and reliability of the rubrics when used an assessment instrument (Taylor, 2011). If these variables are not addressed then the validity and reliability of the assessment and the resulting outcomes becomes questionable (Olfos and Zulantay, 2007).

4. Validity and reliability of rubrics

In the area of education, validity is not seen as a property of the assessment but how the results have been interpreted (Jonsson and Svingby, 2007). Validity refers to the degree to which evidence produced from assessments support the interpretations made about a student's competencies. Table 3 describes the three types of evidence that are commonly examined to support the validity of an assessment instrument: content, criterion, and construct (Moskal and Leydens, 2000).

Validity		
Content Validity: extent to which the assessment instrument provides a representative sample of the content domain in the area of interest (Lynch, 2003)	**Criterion Validity:** extent to which a student's performance on a test accurately predicts the student's performance on an external criterion (Lynch, 2003).	**Construct Validity:** extent to which the assessment measures the theoretical construct or processes that are internal to an individual (Moskal and Leydens, 2000).

Table 3: Three types of evidence commonly examined to support the validity of an assessment

It is extremely difficult to construct an assessment which is truly valid in measuring what it is supposed to measure (Finch, 2002). For example, an assessment designed to assess a student's ability to fight fires may not be able to effectively measure personal or professional behaviours (such as creativity and critical thinking) associated with the task performance. According to Messick (1996), it is hard for assessments to achieve complete validity but he believed that the threats to validity can be minimized by ensuring that assessments do not contain anything that is irrelevant to the measurement of the desired outcomes. For example, assessments designed to assess a student's ability to fight fires should not include pen and paper testing in classrooms which are irrelevant to the measurement of either the task performance or behaviours associated with it.

Does this mean that relevant and authentic scenarios can insure validity?

Capturing a more authentic performance does not insure validity (Stevens, 2013). For example, Hoepfl (2000) pointed out that creating standards for authentic assessments is a challenging task which may suffer from 'Construct underrepresentation' if the standards fail to assess essential dimensions of knowledge and skills or 'Construct-irrelevant variance' if the standards require tasks that are not relevant to measuring the desired competencies (Messick, 1995). Assessments are valid if they effectively measure the intended learning outcome it was designed to assess. Whether assessments effectively measure the intended learning outcomes cannot be based on the subjective judgement of whether questions appear to do so, known as face validity (Drost, 2011). Drost (2011) explains that although face validity is important for credibility to stakeholders, it is the weakest and least scientific form of establishing validity for assessments.

For effective measurement, outcomes should be accompanied by the essential criteria and the levels of performance by which the performance would be judged (Mueller, 2005). The criteria and the levels are usually combined into a rubric, which forms a scoring guide for the assessment making it easier for educators to define what is being measured through assessments and how the score is to be interpreted (Emery, 2001). Scoring without specific guidelines may lead to subjective judgements. Rubrics can be used to improve the objectivity of scoring by specifying the same criteria and standards to be applied to all students' work for scoring by either individual or multiple assessors (Dennison et al., 2015). For example, according to Jonsson and Svingby (2007), one widely cited effect of rubrics in the areas of authentic and performance-based assessments is the consistency of judgement and scoring across students, tasks, and different raters (scorers). Consistency of assessment scores obtained

every time the same competencies are assessed irrespective of the scorer, time period between the assessments, and the context under which the assessments occurred is referred to as reliability (Moskal and Leydens, 2000). Table 4 provides the different types of reliability testing conducted in the area of education.

Reliability			
Inter/Intra Rater: Variations in raters' judgments across raters, known as inter-rater reliability, or in the consistency of one single rater, called intra-rater reliability (Jonsson and Svingby, 2007).	**Test-Retest:** consistency of results when the same test is administered after a specific period (Drost, 2011).	**Split-Half:** Two tests and two measures assessing the same construct (Drost, 2011).	**Internal Consistency:** How well the different components of the assessment measure a particular construct (Drost, 2011).

Table 4: Different types of reliability testing used in student assessments

Ideally, an assessment should produce similar results independent of the scorer and the context of assessment. But is this obtainable?

The more consistent the scores are over different scorers and contexts, the more reliable the assessment is thought to be. Methodologically sound assessment instruments should have acceptable levels of both validity and reliability (Rhodes and Finley, 2013). For example, the study by Vendlinski, Underdahl, Simpson, & Stevens (2002) used rubrics to authentically assess 134 first-year high school chemistry students to achieve valid inferences of a student's content understanding, while not allowing the score to be affected by gender, ethnic, or socioeconomic bias.

The validity of the results and the strength of the rubric as an assessment instrument are evidenced by positive results on a variety of reliability tests (Diller and Phelps, 2008). Performance-based assessments like authentic assessment face the problem of obtaining reliability (Lynch, 2003). Issues such as lack of reliability, inconsistency in assessment design and grading, and potential for grading bias remain important challenges with authentic assessment (Rhodes and Finley, 2013). Authentic assessments represent real-world tasks as valid indicators of workplace competence which should be consistent irrespective of the context or scorer. Such consistency can only be proved through reliability. Hence, authentic assessments should achieve both validity and reliability.

Because it can be difficult to establish whether an assessment instrument truly captures the outcome for which it is intended or whether the outcome can be consistently measured, it is preferable for instruments to demonstrate more than one type of validity (Rhodes and Finley, 2013) and reliability. There are numerous aspects of validity and reliability investigated and reported in the literature on assessment. They may be discussed selectively, but none should be ignored (Jonsson and Svingby, 2007). Although rubrics do not make assessment valid, addressing different aspects empirically could make assessments more valid and reliable for its intended purpose, eliciting the required performance (Jonsson, 2008). There is sparse research focussing on the quality of rubrics as a valid and reliable assessment tool (Stellmack et al., 2009). Hence, a literature review in the area of authentic assessment was carried out to

reveal if a holistic approach to improving its validity and reliability through rubrics has been used by past researchers in the area.

5. Classification of literature

The classification is based on a review of 124 articles which included books, chapters in books, conference papers and proceedings, government documents, journals, reports, thesis, and other articles classified as generic. The articles were chosen after a web-based search on popular websites such as google, google chrome, and google scholar as well the library database of the University of Tasmania. The University of Tasmania uses popular search systems such as ProQuest and Web of Science which enabled to widen the search of articles. Articles were also found by the snowballing technique based on a search through citations in articles discovered through online search. The online search used the phrases 'authentic assessment', authenticity in assessment', and 'authentic+assessment'. Hence, all reviewed articles contain both the words 'authentic' and 'assessment' or 'authenticity' and 'assessment'. The exception being the articles by (Wiggins, 1998) and (BoarerPitchford, 2010). While the former was chosen based on the fact that Wiggins is the most cited author in the area of authentic assessment, the latter was selected due to the discussion of authentic assessment in the research. The articles span from 1989 (when authentic assessment was first introduced) to 2015 (when this paper was being written). An effort was made to obtain as many articles as possible through the above methods.

The purpose of the classification was to highlight the different types of validity and reliability demonstrated in past research, when authentic assessment was implemented with the use of rubrics. As a result, articles where authentic assessment was implemented without the use of rubrics were excluded from the classification. Table 5 provides a snapshot of the criteria used for the inclusion and exclusion of articles from the classification.

Total number of articles selected for the review	**124**
Articles **excluded** based on the non-implementation of authentic assessment (includes theory discussion, theoretical models/frameworks, data collected via interviews; focus groups; and surveys only)	**83**
Articles **excluded** based on implementation of authentic assessment but without the use of rubrics	**24**
Articles **included** based on implementation of authentic assessment with the use of rubrics	**17**

Table 5: The criteria used to select articles for classification

The articles included in the classification were reviewed **(Appendix 1)** to investigate the extent of validity and reliability testing of rubrics in the past, when used as an authentic assessment instrument by researchers for student assessments in various areas of education and training.

6. Gaps found from the literature classification

The intention of the literature classification was to find out the extent of investigation that has been carried out in the area of testing validity and reliability of rubrics as authentic assessment tools. Reliability and validity problems are found to be very typical of authentic assessment (Olfos and Zulantay, 2007). It is often assumed that reliability is achieved concurrently with validity, due to which it may be ignored or accepted with low levels in traditional assessments (Olfos and Zulantay, 2007). This was evident in the study by Olfos and Zulantay (2007) which showed a lack of reliability but showed evidence of validity. So reliability is often accepted as a necessary condition of validity (Olfos and Zulantay, 2007). However, in cases of authentic assessment, reliability cannot be ignored or accepted with low levels as a trade-off between validity and reliability (Jonsson, 2008). Reliability mainly indicates consistency of performance which is essential for workplace-based tasks.

The most obvious gap found in this respect reflects an absence of both validity and reliability testing in some studies such as Todorov and Brousseau (1998), Emery (2001), Vendlinski et al. (2002) and Brawley (2009). Reliability and validity are crucial to the acceptance of authentic assessment as an accurate measure of knowledge, skills, and behaviours (Stevens, 2013). There are numerous extraneous variables that affect the validity and reliability of the rubrics when used an assessment instrument (Taylor, 2011). If these variables are not addressed then the validity and reliability of the assessment and the resulting outcomes becomes questionable (Olfos and Zulantay, 2007). Fook and Sidhu (2010) believe that there is a general lack of research in exploring practices that can improve validity and reliability of assessments through criteria and standards provided in rubrics. The classification reveals that past research in the area of authentic assessment has addressed typically only one or two aspects of validity and reliability while others have not been investigated. The validity was mostly achieved through a review by field experts as evident in the studies by Moon et al. (2005); Fatonah et al. (2013); Olfos and Zulantay (2007); Johnson (2007); Taylor (2011); and Lang II (2012). Barring one study by Jonsson (2008), none of the studies in the classification demonstrated construct validity. A lack of construct validity may indicate that that underlying psychological variables such as problem-solving, social interaction, and communication which are required universally in most professions were not adequately assessed in these cases.

Some studies revealed other types of validity, such as face and convergent validity which were not categorised under the three common types of evidence required to support the validity of an assessment instrument. While face validity is the weakest and least scientific form of establishing validity; convergent validity was explained by Cassidy (2009) as a subcategory of construct validity that seeks “agreement between a theoretical concept and a specific measuring instrument”. The review revealed that some researchers like Cassidy (2009) use a pre-tested instrument expecting the same validity and reliability as obtained in previous studies. However, if using a pre-existing instrument, it is essential for researchers to establish the instrument’s validity and reliability in the context of their own research (Burton and Mazerollw, 2011).

A common method for establishing reliability for rubrics is revealed to be through inter-rater scoring or internal consistency reliability. Reliability in authentic assessments has often demonstrated by a variety of statistical measures and coefficients as evidenced by the studies

of Johnson (2007); Lang II (2012); Olfos and Zulantay (2007); and Diller and Phelps (2008). According to Lovorn and Rezaei (2011), simply using rubrics do not improve the reliability of the assessment. Reliability can only be improved if rubric users are well-trained on its' development and use. Raters/Scorers need to be involved in the development of rubrics or else it takes time for them to understand its purpose and implementation (Diller and Phelps, 2008). For example, the study by Lovorn and Rezaei (2011) involved the training of 55 teachers in rubric use to find a resulting increase of reliability in writing assignments. However, many of the studies such as Moon et al. (2005), Olfos and Zulantay (2007), Diller and Phelps (2008), do not mention any training for rubric users before they were administered. In the study by Taylor (2011), teacher development workshops were carried out to minimize threats to internal validity only. However, according to Taylor (2011), training conducted for rubrics development or use should be consistent for all involved. Differing approaches in terms of context, standards, or application can impact the results of research data and create problems with validity.

The classification also reveals an absence of research of authentic assessment in the field of seafarer education and training. Past research (Bell and Bell, 2003; Cassidy, 2009; Wellington, Thomas, Powell, and Clarke, 2002) showed that authentic assessment has been implemented to investigate its impact on achievement of educational or professional standards, constructive alignment of instruction processes with assessment, and achievement of professional competence (including demonstration of essential behaviours). Similar research is needed but has been largely ignored in the area of seafarer education.

7. Conclusion

The move of the STCW'95 code towards OBE highlights the need of assessment practices that allow demonstration of learning outcomes by seafarer students through performances in real-world or contextually similar settings provided by authentic assessment.

To validate if intended outcomes are being measured consistently through assessments, authentic assessments need to achieve validity and reliability through clear statements of learning expectations provided by assessment rubrics. The validity and reliability of the rubric is not only essential for the validation of outcomes attainment but also for the rubric to be accepted as an instrument of authentic assessment that can effectively measure outcomes. An extensive literature review in the area of authentic assessment revealed a lack of research in a holistic approach to addressing different aspects of validity and reliability of rubrics when used as an authentic assessment instrument. The absence of a robust framework challenges and undermines the resulting outcomes from the learning and teaching experience attained by past researchers who based their findings using rubrics that addressed only selected aspects of validity and reliability. While addressing different aspects of validity will identify and assess the content and essential underlying constructs of professional competence in different contextual scenarios; different aspects of reliability will assure consistency in performance. Overall, this will ensure a holistic approach to competence assessment at a standard expected in employment.

Past research provides theoretical justification and empirical evidence of the value of authentic assessment when educators are seeking to:

1) obtain evidence of the development and achievement of professional competence;

2) raise the standards of student performance and achievement;

3) measure the effectiveness of the teaching and learning

4) develop higher-order and critical thinking skills in students; and

5) successfully align learning, teaching, and instruction with assessment.

The above outcomes together with a holistic approach to competence assessment will also benefit seafarer education and training. While knowledge-based components may continue to be assessed via traditional examinations, application of skills in real-world contexts will engage seafarer students through meaningful and relevant learning. Authentic assessments will go beyond meaningful contexts and also require seafarer students to integrate competence acquired for different STCW tasks for a holistic workplace-based performance. For example, assessment for the STCW task of 'planning and conducting a passage and determine position' may be designed to integrate components from other STCW tasks such as, 'maintain a safe navigational watch', 'use of ECDIS to maintain the safety of navigation', and 'manoeuvre the ship'. Assimilating, analysing, and integrating information from different units of competence will make the seafarers active participants in the process of learning and enhance student engagement. Demonstrating competence in authentic contexts will provide seafarer students with an understanding of how skills acquired in classrooms may be transferred at the workplace. Using pre-established performance criteria, students will frequently reflect on their current level of learning and compare it with the level required at the workplace, allowing them to develop strategies for raising their standards of performance.

The review reveals that there is a lack of global research on authentic assessment in the field of seafarer education and training. Further research needs to establish how to use authentic assessment within the confines of the STCW Code to improve:

1) student engagement;

2) transfer of competence; and

3) standards of performance.

Inherent to such future research, investigations shall also reveal ways to:

1) increase the validity and reliability of rubrics as an authentic assessment instrument; and

2) use rubrics as an authentic assessment instrument to satisfy employer and regulator expectations with the attainment of the standards stipulated in the STCW Code.

7.5. Paper V

IMPROVING THE VALIDITY AND RELIABILITY OF AUTHENTIC ASSESSMENT IN SEAFARER EDUCATION AND TRAINING: A CONCEPTUAL AND PRACTICAL FRAMEWORK TO ENHANCE RESULTING ASSESSMENT OUTCOMES

Samrat Ghosh
Australian Maritime College, University of Tasmania
Marcus Bowles
Future of Professional Work and Learning, Deakin University
Dev Ranmuthugala
Australian Maritime College, University of Tasmania
Ben Brooks
Australian Maritime College, University of Tasmania

Abstract

Past literature on authentic assessment suggests it provides a far more reliable and valid indicator of outcomes such as higher student engagement, ability to transfer skills to different contexts, multiple evidence of competence, and student performance. This has appeal in seafarer education and training where both students and employers increasingly perceive traditional assessment methods as failing to consistently generate these outcomes. However, this paper argues that improving different aspects of assessment validity and reliability is essentially required to enhance the outcomes identified above. The paper builds on and extends previous work to investigate and develop a conceptual and practical framework that promotes a holistic approach to authentic assessment that provides greater assurances of validity and reliability throughout all stages of assessment within seafarer programs. It also lays the path to future research directions by establishing the agenda to test the practicality of the framework in the authentic assessment of seafarer students and also investigate the impact of students' perception of increasing authenticity on performance scores in assessment tasks.

Keywords – Authentic assessment, validity, reliability, seafarer, education and training

1. Introduction

The STCW Code provides vague descriptions of standards of competence, where each standard is discrete and does not necessarily require holistic assessment (Ghosh, Bowles, Ranmuthugala, and Brooks 2014). Training and assessment standards need to exceed the minimum STCW requirements (AGCS 2015) to ensure operational errors causing expensive maritime disasters are reduced to a minimum.

The assessment tasks should ideally assess the students' ability to perform workplace tasks at standards required on board ships. This was recognised by IMO when revising STCW'78, which was essentially knowledge-based comprising a syllabus for qualifying examinations instead of focusing on skills and abilities necessary to perform workplace tasks (Morrison 1997). STCW'78 was revised to create STCW'95, which essentially required seafarer students to demonstrate their competence to standards prescribed in the Code. Although, many of the MET providers use simulators and practical exercises for training and assessment in selected units of the STCW Code, the use of decontextualized traditional assessment methods (e.g. multiple-choice questions, pen and paper testing, oral examinations) for most of the units of competence listed in the STCW Code cannot be ignored. Past research (Emad and Roth 2007; Cox 2009; Sampson, Gekara, and Bloor 2011) showed that seafarer students and employers perceive decontextualized traditional assessments to be falling short in their ability to replicate workplace settings and as a result to: engage students and develop their ability to transfer learning to different contexts.

For example, an ethnographic case study involving 16 students carried out by Emad and Roth (2007) in a Canadian maritime institute, revealed that students disengaged with traditional exams that comprised mainly of the questions which were drawn from a question bank. Similarly, a study by Sampson et al. (2011) revealed that employers were unhappy with some of the current assessment methods that assessed a limited range of job specific skills, in settings that provide insufficient cues to the students on how the competence acquired in classrooms are applied in different contexts. Official investigations and analysis of marine accidents have also revealed that seafarers assessed as competent in the use of particular skills in a given context, failed to apply them in others (Uchida 2004; Pecota and Buckley 2009; Prasad, Nakazawa, and Baldauf 2010).

More effective educational practices will enhance student performance and also meet stakeholder expectations (McLaughlin 2015). The expectations of the students and employers may be addressed if Seafarer Education and Training (SET) implements authentic assessment that require students to emulate task performance at workplace standards in real-world contexts (Bosco and Ferns 2014). However, to ensure authentic assessment has a high fidelity to real-world contexts and requires competence as expected at the workplace, it should be judged for its technical adequacy of measures by the established criteria of validity and reliability (Linn, Baker, and Dunbar 1991). Addressing different aspects of validity and reliability will not only provide evidence of a student's ability to perform assessment tasks using real-world competencies and to workplace standards but also to do so consistently, ensuring a holistic approach to competence assessment. Hence, this paper addresses the following objectives:

• Establish theoretically how addressing different aspects of validity and reliability of authentic assessment will lead to higher student engagement, student ability to transfer skills to different contexts, contextual and multiple evidence of competence, and valid and reliable student performance.

• Investigate the existence of a framework that has a holistic approach to the validity and reliability testing of authentic assessment based on an extensive literature review of 152 articles.

• Construct a conceptual and practical framework that addresses and improves upon the different aspects of validity and reliability of authentic assessment during specific stages of its implementation.

The framework will also generate the future research agenda of testing and operationalising to investigate the impact of students' increasing perception of authenticity in assessment on performance scores in tasks.

2. Authentic assessment needs a holistic approach to validity and reliability

Traditional assessments that focus on written or oral examination of knowledge may be effective in assessing students' ability to memorise and regurgitate knowledge-based components of the task. However, they are poor foundations to determine demonstrated skills, deep understanding or overall outcomes from learning unless they are integrated with performance-based assessments, such as authentic assessment, to reflect attainment of standards expected in the workplace (Biggs and Tang 2010; O'Farrell 2005). An extensive literature review of 124 articles in the area of authentic assessment (presented previously in Ghosh et al. 2015) defined it by collating the characteristics highlighted by the more highly cited authors (e.g. Wiggins 1989; Archbald 1991; Darling-Hammond & Snyder 2000). According to the characteristics collated, authentic assessment will encompass three aspects: tasks, processes, and outcomes as presented in Table 1.

Authentic Assessment		
Tasks	**Processes**	**Outcomes**
• Set in a real-world context • Requiring an integration of competence • Comprising of forward looking questions • Ill-structured problems	• Requiring performance criteria to be provided beforehand • Evidence of competence to be collected by the student	Resulting in: • Higher student engagement • Ability to transfer skills to different contexts • Contextual and multiple evidence of competence • Valid and reliable student performance

Table 1: Definition of authentic assessment based on characteristics provided by the more highly cited authors

According to the definition, the tasks and processes of authentic assessment should result in the outcomes of: higher student engagement, ability to transfer skills to different contexts, contextual and multiple evidence of competence, and valid and reliable student performance. Since traditional assessments and those linked to the STCW Code frequently fall short in their ability to achieve these outcomes within SET, the implementation of authentic assessment may provide the tools to address the perception of seafarer students and employers with regard to these shortcomings. However, to ensure that the 'authentic' tasks reflect workplace situations requiring students to apply knowledge, skills, and behaviours to professional standards; and to test consistency of such performances, authentic assessments and the resulting performances should be judged by the essential criteria of validity and reliability. Validity and reliability in assessments is not a property of the assessment but the interpretation and consequences of assessment scores (Messick, 1995; 1996).

In the evaluation of the quality of student assessments, validity refers to the degree to which evidence produced from assessments support the interpretations made about a student's competencies; and reliability can be defined as the degree of the consistency of assessment scores obtained every time the same competencies are assessed irrespective of the scorer, time period between the assessments, and the context under which the assessments occurred (Moskal and Leydens, 2000). The different types of validity for performance-based assessments comprise of content, criterion, and construct validity (Messick 1995; 1996; Linn et al. 1991). The different types of reliability for assessments include test-retest, split-half, internal consistency (McAlpine 2002) and inter-rater (Jonsson 2008) reliability. The different types of validity and reliability are essential for authentic assessment (tasks and processes) to achieve its intended outcomes of:

- higher student engagement;
- ability to transfer skills to different contexts;
- contextual and multiple evidence of competence; and
- valid and reliable student performance.

The following sections discuss how improving the validity and reliability of authentic assessment will contribute to the achievement of the four outcomes listed above.

2.1. Higher student engagement

Authentic assessment requires tasks to resemble real-world scenarios or similar contexts. Real-world scenarios provide meaningful contexts for knowledge and skill application to students, thus, creating a high level of student engagement and commitment (Richards Perry 2011; Pallis and Ng 2011). However, how do we ensure that the authentic tasks designed by the educators are perceived by the seafarer students as valid and relevant to workplace tasks?

Content validity evaluates the extent to which the assessment instrument provides a representative sample of the content domain in the area of interest (Lynch 2003). For example, if the authentic assessment was designed to assess a seafarer student's competence to fight

fires on board a ship, content validity of the assessment will ensure that it adequately covers the content of fire-fighting practices and conditions on ships. It will also ensure that the assessment does not contain anything that is irrelevant to the measurement of the ability to fight fires. Hence, content validity is popularly achieved through validation by subject experts (Oh, Kim, Garcia, and Krilowicz 2005; Lang II 2012). However, it is a rational analysis based upon individual, subjective judgment (Moon, Brighton, Callahan, and Robinson 2005), which may result in bias. The bias may be reduced if multiple subject experts are employed for validation. (Moon et al. 2005). To be engaged in learning, students will not only require meaningful contexts but also to be active participants in the knowledge construction process (Hart, Hammer, Collins, and Chardon 2011). According to the learning theory of constructivism, construction of knowledge allows students to develop a deeper understanding of the learning content (Biggs 1999). Authentic pedagogical practices are influenced by the constructivist philosophy of student-centred learning, where students create meaningful knowledge in real-world tasks (Morrissey 2014), thus engaging students in the learning process (Quartuch 2011). The question is how do we ensure that the authentic tasks require seafarer students to construct knowledge using competencies (technical and soft/underpinning skills) as required in the real-world?

Construct validity evaluates the extent to which the assessment measures the theoretical construct or processes that are internal to an individual (Moskal and Leydens 2000). For example, construct validity will ensure that the authentic assessment of a student's ability to fight fires on board a ship not only assesses the technical knowledge of fire-fighting but also the essential and critical underpinning/soft skills of problem solving, communication, and critical thinking. The development of 'soft' skills in students allows them to transfer these skills into different scenarios and roles/responsibilities (Mitchell 2008) and may also create higher student engagement. The recognition of the soft skills and the requirement to assess them is essentially missing within the STCW Code (Ghosh et al. 2014).

Student engagement may be higher if students are provided with clear expectations of learning standards to be achieved before the assessment is implemented (Findlay 2013). Students are then measured against identified standards of achievement, and how well the individual student has performed by applying specific criteria and standards (Dunn, Parry and Morgan 2002). Standards are defined as levels of definite attainment and sets of qualities established by authority, custom, or consensus by which student performance is judged, whereas criteria are essential attributes or rules used for judging the completeness and quality of standards (Sadler 2005; Spady 1994). Although, such criterion referenced assessments are promoted in the performance-based assessments like authentic assessment, traditional assessments shy away from doing so and follow the norm referenced assessments (Dikli 2003). Hence, norm referenced assessments that do not inform students on standards of achievement, if implemented in SET, will not assure minimum competence (Lister 2006).

Providing students with essential criteria and standards of achievement at the beginning of the learning period is an essential requirement of the authentic assessment process (Wiggins 1989; Archbald 1991; Darling-Hammond & Snyder 2000). In authentic assessment, the teacher provides a roadmap of the entire subject to be learned while allowing students to construct their understanding of the topic. Providing standards of performance beforehand enables students to reflect on their learning and carry out self-assessments of their thinking and

practices towards achievement of the required standards (Findlay 2013). As learning progresses, learners assume increasingly more control over the sequence in which they want to engage their learning (Schell, 2000) and gain mastery over knowledge and skills learnt through strategic and critical thinking (Fredricks and McColskey 2012). Seafarer students are expected to achieve learning outcomes driven by the STCW Code. However, lack of descriptive outcomes within the Code (Ghosh et al. 2014) and traditional teaching and assessment practices often do not provide the students with clear expectations of the learning standards to be achieved.

The use of assessment rubrics is one method of providing the students in advance the performance criteria and standards to be achieved (as required in authentic assessment) as well as adhering to the competency standards (Diller and Phelps 2008) such as the STCW Code in SET. Rubrics are assessment tools that comprise of individual and essential dimensions of performance known as criteria along with standards for levels of performance against those criteria (Jonsson and Svingby 2007). Using the objective standards and criteria, assessment rubrics can be used for evaluating student performance and providing them with feedback on the level of learning achieved (Diller and Phelps 2008).

Providing feedback on student performance allows educators to identify areas of learning that need improvement. Hence, assessment rubrics can be a very effective tool to obtain inter/intra-rater (scorer or assessor) reliability. Inter-rater reliability evaluates the variations in judgments across raters, while intra-rater reliability looks at the consistency of one single rater (Jonsson and Svingby 2007). The assessment rubrics can be used as a common marking guide by the raters, where the objective standards and criteria may promote unbiased marking (Oh et al. 2005). However, to obtain a high inter-rater reliability, rigorous training of raters may be essential to avoid differing approaches to marking (Koh and Luke 2009; Taylor 2011). Ideally raters should be involved in the development of assessment rubrics, otherwise it will require time and effort to ensure they understand its purpose and implementation (Diller and Phelps 2008).

On completion of rater marking, assessment rubrics may be used to provide students with feedback on standards of learning achieved. The feedback may be used by students to engage in meaningful reflection, known as metacognition (Scott 2000). Students reflect on their current level of learning and engage in self-assessment which allows them to identify the gaps between their current competence and those required by educators or employers at the workplace (Boud and Walker 1998). Recognizing gaps in their knowledge allows students to develop strategies towards filling those gaps making learning more structured and deeper. This is a departure from the 'surface' learning approaches that students engage in for purely passing examinations, and hence, may engage students in learning. Ability to recognize gaps in knowledge through self-assessment also develops students' understanding of how skills developed in particular contexts may be used in different contexts. This will enable seafarer students to understand key requirement for the transfer of learning from the classroom context to ships as a workplace (McCarthy 2013).

2.2. Ability to transfer skills to different contexts

Students who are able to frequently reflect on their learning to recognize gaps in their own construction of knowledge and improve on them, begin to grasp cues (Leberman 1999; Sator 2000) on applying the same knowledge (developed in a specific context) to different contexts resulting in transfer of learning (Bransford, Brown, and Cocking 2000; Donovan, Bransford, and Pellegrino 1999). Students re-evaluating their learning, develop critical thinking skills causing behavioural changes that promote positive growth in cognitive development, which can be used to assimilate, analyse, and structure information for decision making and problem solving (Saunders, Saunders, and Batson, 2001). Cognitive development provides students with the belief and confidence (Bandura 1977) to transfer newly acquired knowledge and skills (Merriam and Leahy 2005). Learners draw on and extend previously learned knowledge and develop their own cognitive maps to interconnect facts, concepts and principles. As learning progresses, understanding becomes integrated and structured, leading students to gain mastery over content (Scott 2000). Students' ability to transfer is enhanced when they are able to use the deep understanding of the learning content to interconnect facts and apply it to different contexts (Mestre 2002).

However, according to the constructivism theory of learning, transfer is enhanced when learning is contextualized in authentic tasks designed in meaningful contexts (Ertmer and Newby 1993). Providing authentic tasks that require application of knowledge as in the real-world, will allow students to identify essential 'threshold concepts' central to facilitate transfer of learning (Moore 2012). The authentic tasks which may initially be unfamiliar to students will comprise of cues to facilitate understanding of transfer. The cues allow students to gain an understanding of threshold concepts required to master the subject and understand how it may be integrated with other units of learning (Cousin 2006). As the complexity of the tasks is increased, fewer cues are provided for students to practice transfer of learning in dissimilar situations.

Due to the complexity in recreating the shipboard workplace environment in land-based MET institutions, most of the learning and assessment in seafarer education takes place in decontextualized scenarios. Herrington and Herrington (1998) indicate authentic assessments conducted in real-world contexts will provide 'cues' to students on strategies to adopt when performing in the real world. Contextualised authentic tasks may not recreate all of the conditions within a shipboard workplace but may replicate many of the complexities and challenges faced by seafarers in the real-world. Content and construct validity may ensure that the assessment tasks resemble real-world scenarios requiring the targeted competencies in order to perform adequately within that environment to the required workplace standards. However, capturing a more authentic performance does not ensure validity (Stevens 2013). Testing for internal consistency reliability may be one of the ways of avoiding this problem.

Internal consistency evaluates how well the different components of the assessment measure a particular construct (Drost 2011). Internal consistency measures 'consistency' within the assessment instrument and based on the average inter-correlations among all the single items within the test, questions how well the items measures (Drost 2011) particular learning outcomes and/or behaviours associated with the learning outcome. Internal consistency reliability can be measured via various statistical measures (Oh et al. 2005; Olfos and Zulantay

2007; Cassidy 2009) and some of these methods (split-half and test-retest reliability) may also generate multiple evidence of competence.

2.3. Ability to transfer skills to different contexts

Internal consistency reliability can be measured using statistical measures such as Kuder Richardson #20 (Jonsson 2008) or Cronbach's coefficient alpha (Oh et al. 2005) which determine the correlations of the test questions to the competency it purports to measure. This may also be done using the split-half or test-retest reliability (Drost 2011). Split-half reliability involves administering two separate tests or splitting an individual test to create two measures (result of one half compared with the other) assessing the same construct (Drost 2011). However, irrespective if it is one single test or two separate tests, all questions should measure the same construct (McLeod 2013). Test-retest reliability involves administering the same test after a specific period of time (Drost 2011). Timing of the test becomes an important variable in this type of reliability test. If the duration between the tests is too short, the students may recall information from their first attempt which may bias the result. Alternatively, if the duration is too long, there may be a 'learning effect' due to extraneous variables that may not be easily identified (McLeod 2013). In either case of split-half or test-retest reliability, the statistical measures of correlation between test questions provide internal consistency reliability. Additionally, students assessed on two separate tests or the same test twice will not only evaluate consistency in performance but also provide multiple evidence of competence and will confirm the students' ability to repeat performance.

Multiple evidence of competence may also be generated if the assessment is tested for criterion validity. Criterion-related validity evaluates the extent to which student scores on an assessment relate to scores on a previously established but valid assessment implemented approximately simultaneously (concurrent validity) or in the future to a measure of some other criterion that is available at a future point in time (predictive validity) (Lang II 2012). The administration of multiple assessments should also be followed by inter-rater reliability where two or more raters evaluate the student work. The use of assessment rubrics in this case will not only provide evidence of achievement against the learning standards and criteria but also act as a contextual evidence of competence. The rubrics along with the standards and criteria may also detail the context under which the task was performed, and competence acquired. Multiple evidence of competence may enhance the seafarer employer's perception of the quality of evidence produced via authentic assessment. If the evidence demonstrates the seafarer student's ability to perform authentic tasks that represent real-world scenarios requiring competencies as required at the workplace; and to do so repeatedly and consistently (as verified by multiple raters), seafarer employers may perceive the assessment and the resulting performance to be more valid and reliable.

2.4. Valid and reliable student performance

Authentic assessment should not only assess the seafarer students' ability to perform real-world tasks to workplace standards (valid performance) but the ability to do so consistently (reliable performance). The student performance in the assessment tasks should allow valid generalizations about student competence (Wiggins 1992) with respect to the demonstrated learning outcome. However, such generalization cannot be based on one performance, irrespective how complex or authentic the task was (Wiggins 1998). One of the ways generalizability across tasks may be achieved is to increase the number of performance assessments for each student providing them with more than one opportunity to demonstrate their mastery over the competence (Linn, Baker and Dunbar 1991). Criterion-related validity or split-half reliability (using two separate tests) of authentic assessments will provide students with more than one opportunity to demonstrate mastery over the construct that will be commonly measured in the tests implemented.

Data derived from valid and reliable student performances may be used to identify ways to improve the different aspects of validity and reliability of authentic assessment, which in turn may enhance student performance of tasks. For example, Jonsson (2008) found the overall student scores increased by over 60 percent when transparency of rubrics was increased based on student performances in the previous year. This example shows that although authentic assessment does not assure enhanced student performance, its validity and reliability testing will provide evidence towards change in teaching practices that may result in improved performance.

The above discussion reveals that authentic assessment will be able to achieve its intended outcomes if it addresses and improves upon the different aspects of its validity and reliability. In the context of SET, if the numerous extraneous variables that affect the validity (content, construct, and criterion) and reliability (inter-rater, internal consistency, split-half, and test-retest) of the authentic assessment are not improved, then the resulting evidence of competence may become questionable (Olfos and Zulantay 2007) to seafarer employers, adversely affecting the employment of seafarer graduates and defeating one of the key purpose of their education and training. Hence, there is need for a framework that has a holistic approach to the validity and reliability of authentic assessment.

3. Investigating the need for a conceptual framework

This section builds on and extends the previous literature review conducted by the authors (Ghosh et al. 2015) with an aim to investigate the existence of a conceptual framework that has a holistic approach to validity and reliability of authentic assessment.

The review for this section included using the title and abstract search of the library (University of Tasmania) and google databases with the following keywords and Boolean operators:

"authentic assessment" OR "authenticity in assessment" OR "authentic" OR "authenticity" OR "performance assessment"

AND

"seafarer education and training" OR "engagement" OR "transfer" OR "validity" OR "reliability" OR "evidence of competence" OR "rubrics" OR "student performance".

The first set of keywords reflect those used in the main literature review conducted in the field of authentic assessment by past researchers (Ashford-Rowe 2009; Taylor 2011; Varley 2008). The second set of keywords were used to identify published research that investigated the relationship between authentic assessment and the outcomes of engagement, transfer of learning, evidence of competence, and valid and reliable student performance outcomes. In comparison to the previous work by the authors that reviewed 124 articles (Ghosh et al. 2015), this review obtained and reviewed a total of 152 articles (from 1989 when authentic assessment was first introduced to 2016).

The review analysed the 152 articles where authentic assessment was implemented for student assessment, and the extent of validity and reliability testing conducted on the assessment. Of the 152 articles, 49 articles were based on the implementation of authentic assessment of student learning. Only 12 of those 49 articles addressed one or two aspects of validity and reliability. The remaining 37 articles implemented authentic assessment for students but did not address any of the aspects of its validity and reliability. The analysis of the 12 articles failed to reveal any existing conceptual frameworks that addressed the different aspects of validity and reliability associated with authentic assessment. Table 2 summarises the analysis of the 12 articles.

Author (Year)	Level of studies applied	Validity Tested	Reliability tested	Conceptual framework for validity and reliability testing
Moon et al. (2005)	Secondary school students	Content	Inter-rater	None
Oh et al. (2005)	Undergraduate university biomedical science students	Content	Inter-rater; Internal consistency	None
Johnson (2007)	Secondary and high school students	Face; Content	Internal consistency	None
Olfos and Zulantay (2007)	Primary and secondary school teacher students	Criterion	Internal consistency	None
Jonsson (2008)	Undergraduate university teacher students	Face; Construct	Inter-rater; Internal consistency	None
Diller and Phelps (2008)	Undergraduate university librarian students	None	Internal consistency	None
Cassidy (2009)	Elementary school teachers	Construct	None	None
Koh and Luke (2009)	Elementary and high school teachers and students	None	Inter-rater	None
Taylor (2011)	Secondary and high school students	None	Inter-rater	None
Gao and Grisham-Brown (2011)	Elementary school students	Criterion	None	None
Fatonah et al. (2013)	Elementary school students	Content	Inter-rater	None
Hensel and Stanley (2014)	Undergraduate university nursing students	None	Inter-rater	None

Table 2: Absence of existing conceptual framework towards improving validity and reliability of authentic assessment holistically

Table 2 introduces the use of face validity by researchers like Johnson (2007) and Jonsson (2008). Face validity is achieved through the subjective judgement of experts on the suitability of the content of the assessment towards the measurement of a particular construct (Secolsky, 1987). Since face validity is based on subjective judgement on what may "appear" to be a good measure, it is considered to be the weakest and least scientific form of establishing validity (Drost, 2011).

Table 2 reveals a global absence of a framework that identifies and practically improves upon the different aspects of validity and reliability of authentic assessment, justifying the need to develop one especially in the context of SET.

4. A conceptual and practical framework for improving validity and reliability of authentic assessment in SET

The conceptual framework developed in this paper identifies and improves upon the different aspects of validity and reliability at different stages of the authentic assessment implementation. Based on the definitions of the different aspects of validity and reliability discussed in this paper and their uses in the past research, the development of the framework is discussed in the three specific stages of:

• Before the implementation of authentic assessment;

• During the implementation of authentic assessment; and

• After the implementation of authentic assessment.

4.1. Before the implementation of authentic assessment

It is a requirement of authentic assessment to design tasks in a real-world context. Hence, the focus of authentic assessments for validity purposes should be on creating tasks that emulate workplace challenges faced by practicing professionals. Therefore, it is critical that before authentic assessment is implemented, the designed task should be tested against the desired workplace standards to assure content validity (Moon et al. 2005) and construct validity (Wiggins 1998). Content validity should ascertain if the authentic tasks resemble real-world scenarios, encompassing wide but required content and assessing only intended outcomes. Thus, content validity is generally attained through a review by subject experts. Construct validity should ascertain if the task performed required an integration of competence acquired in individual units of learning, using not only technical/occupational skills but also the essential soft/underlying skills. It should also ensure that the tasks comprise of forward-looking questions and ill-structured problems as required in authentic assessments. Jonsson (2008) explained that construct validity can also be achieved through subject experts' validation before the authentic assessment is implemented.

The performance criteria should be provided beforehand and at the beginning of the learning period to the students. This should preferably be carried out through assessment rubrics as they detail the essential criteria and standards to be achieved by the students. Providing assessment rubrics beforehand will allow the students to use it as a guide before as well as during assessments, to develop strategies towards the collection of evidence required to demonstrate competence at the required standards of learning.

4.2. During the implementation of authentic assessment

Once the authentic assessment is implemented, the student performance should be marked using the inter-rater reliability approach. Inter-rater will use more than one rater (scorer) to ascertain the consistency of the results obtained. The assessment rubric is useful to the scorers as it provides them with clear guidelines on the essential criteria and standards of performance expected from the students. Using the same rubric for assessment and marking will inject objectivity and fairness in the results obtained. In evaluating scores involving raters, it is

important to know the extent to which different scorers agree (or disagree) on the values assigned to student responses (Moon et al. 2005). Cases where multiple raters do not agree with the values assigned to student performance may produce a discrepancy in the resulting evidence of competence and create employer dissatisfaction. Hence, to establish more consistency and reliability in scoring, the framework may need to adopt a practical approach of using a two-member rater panel with a third panel member available for arbitration in case of a disagreement between the raters (Taylor 2011).

4.3. After the implementation of authentic assessment

Once authentic assessment is implemented, and the initial evidence of competence is acquired, the framework should establish the internal consistency reliability to determine the degree to which individual items that comprised the assessment, consistently measure the same objectives. Finally, the framework should employ criterion validity to compare the effectiveness of the authentic assessment task to measure the professional competence with a secondary assessment. The secondary assessment should be an existing but valid assessment that measures the same construct (Gao and Grisham-Brown 2011) and may be implemented concurrently or at a later date. Employing concurrent validity will generate multiple evidence of competence to perform the task and the students' ability to use the underlying competencies.

The effectiveness of the framework to address validity and reliability of authentic assessment; and of its ability to generate the stipulated outcomes is verified via a feedback loop provided at the end of the framework. This is because the effectiveness of the valid and reliable authentic assessment of students is ascertained only after the event. Data from student performances will provide valuable inputs towards the improvement of the assessments. While student and employer perceptions will provide feedback on the authentic assessment outcomes, data from student performances will provide the necessary feedback to enhance the validity and reliability of authentic assessments. Once the feedback is obtained, the loop takes the educators back to the design stage of the assessment task. Modifications based on the feedback will enhance the validity and reliability of authentic assessments; and in turn the resulting outcomes of assessment.

Figure 1 describes the conceptual framework created for the purpose of improving the validity and reliability of authentic assessment when implemented for seafarer education and training. The authentic assessment framework for SET (AAFSET) employs a holistic approach to the validity and reliability of the authentic assessment. However, the framework is conceptual in nature and needs to be tested. The next section details the research that needs to be conducted to test the framework developed in this paper.

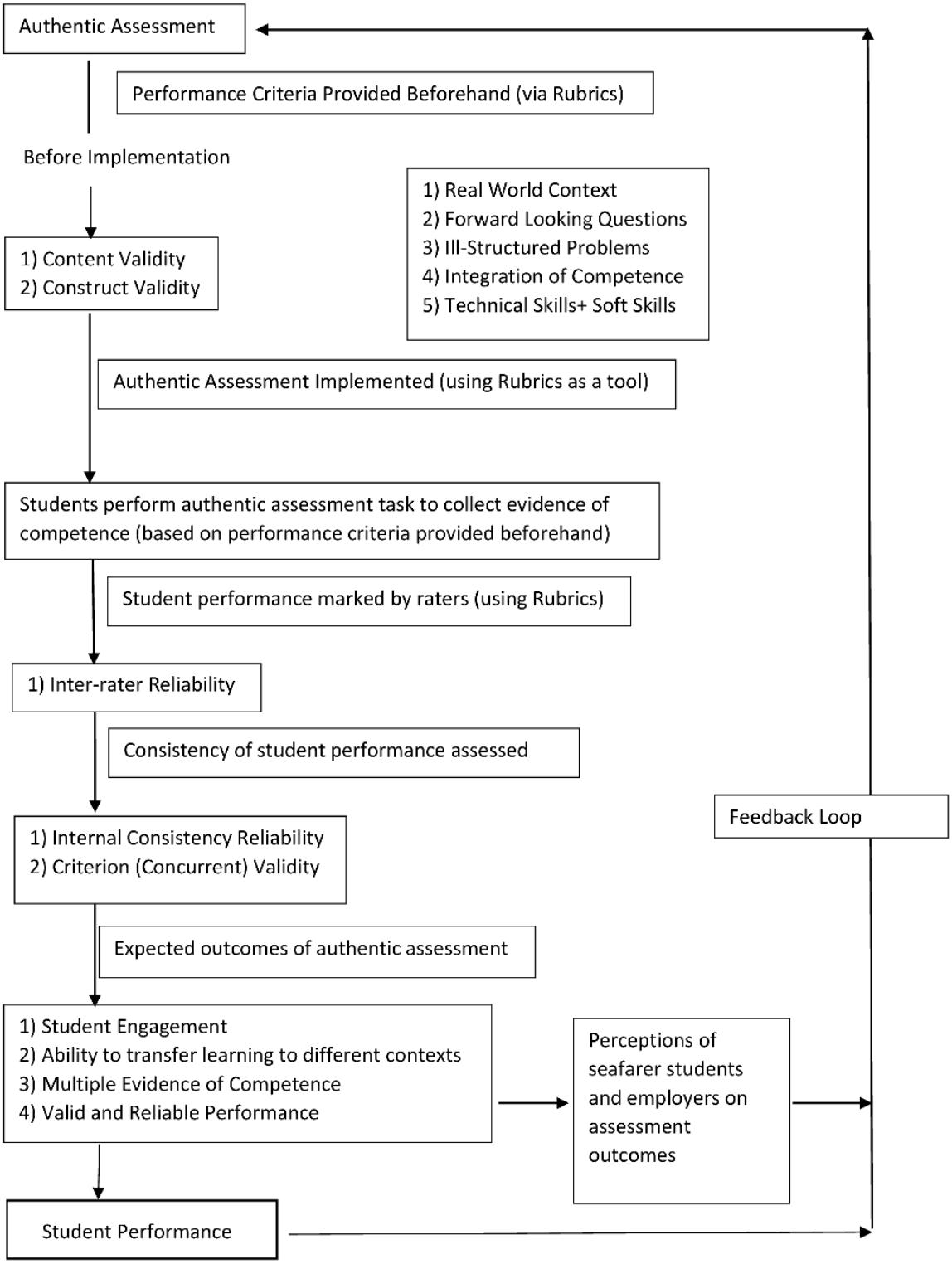


Figure 1: Authentic Assessment Framework for Seafarer Education and Training (AAFSET)

5. Future research: The Way Forward

Based on empirical evidence and theoretical reasoning, this paper argued that validity and reliability testing in authentic assessment will enhance its authenticity and the resulting student performance should provide a valid and reliable indicator of their competence to perform similar tasks at the workplace. Future research shall investigate the impact of authenticity in assessment on student scores in task performances, especially in the context of SET. To do so, a comparative study between traditional, and valid and reliable authentic assessment will be conducted through implementation in a unit of learning that forms part of seafarer certification. The research shall take place over two semesters and use a common unit of learning that is offered in both semesters. While traditional assessment will be implemented for the students enrolled in the first semester, valid and reliable authentic assessment will be implemented for a separate cohort of students enrolled in the second semester. The research study shall investigate how student perceptions of increasing authenticity (from traditional to authentic assessment) impact their performance and resulting scores.

6. Conclusion

As technologies, practice and compliance standards enforced by nations change, seafarer students and employers perceive current assessment methods employed by MET institutes to be deficient in terms of four essential outcomes: student engagement, ability to transfer from the learning to diverse workplace contexts, contextual and multiple evidence of competence, and valid and reliable student performance. This paper examined empirical evidence and theories from the literature to identify authentic assessment as a possible solution to address these expectations and perceptions. It argues that validity and reliability are essential technical measures for evaluating the quality of authentic assessment; and the various aspects of validity and reliability need to be improved to achieve the intended outcomes of the assessment. An extensive review of literature in the area of authentic assessment revealed an absence of an accepted framework that describes a systematic and holistic approach to improving validity and reliability through the use of authentic assessment. Building on existing research this paper makes a theoretical contribution in the area of authentic assessment via a hypothesized relationship, that is ‘if aspects of validity and reliability of authentic assessment are improved holistically, then assessment of SET and the resulting evidence of student competence to perform workplace tasks can be significantly improved’. It will crucially, raise the positive perceptions of students and employers with regard to the resulting assessment outcomes, assuring that the assessment is to a standard they can ‘trust’. Based on the hypothesized relationship, this paper makes a methodological contribution by developing a conceptual framework to address and improve the various aspects of validity (content, construct, and criterion) and reliability (internal consistency, inter-rater, split-half, and test-retest) during the different stages (before, during, and after) of the implementation of authentic assessment. The framework that incorporates a feedback loop will use valuable data from student performances, and student and employer perceptions to enhance validity and reliability of authentic assessment and its resulting outcomes. Although this paper is conceptual in nature, it provides the foundation for future research where the framework will be tested for its’ practicality of use in the authentic assessment of seafarers. Further research is also required to

investigate the impact of seafarer students' increasing perceptions of authenticity on their performance scores in the assessment task.

7.6. Paper VI

AUTHENTIC VERSUS TRADITIONAL ASSESSMENT: AN EMPIRICAL STUDY INVESTIGATING THE DIFFERENCE IN SEAFARER STUDENTS' ACADEMIC ACHIEVEMENT

Samrat Ghosh
Australian Maritime College, University of Tasmania
Benjamin Brooks
Australian Maritime College, University of Tasmania
Dev Ranmuthugala
Australian Department of Defence
Marcus Bowles
Deakin University

Abstract

Past research showed that traditional assessments that required seafarer students to focus on rote learning and construction of responses devoid of context (e.g. oral examinations and multiple-choice questions) disengaged them from learning. Memorising information is a lower-order cognitive ability, failure in which led to errors in assessment tasks resulting in low academic achievement for students. Authentic assessment presents a model that requires students to construct responses through the critical analysis of information presented in real-world contexts. Hence, this research project investigated the difference in seafarer students' academic achievement (measured through scores obtained in assessment) in authentic assessment as compared with traditional assessment. Two separate and independent student groups as the 'control' and 'treatment' group were used for a selected unit of learning delivered at the Australian Maritime College within the Bachelor of Nautical Science degree program. Since, some past researchers defined and implemented traditional assessments as a single-occasion assessment, this project implemented the assessment in a summative format as opposed to authentic assessments implemented formatively. Analysis of student scores revealed that the authentically assessed students were guided towards significantly higher academic achievement. Project findings identified vital challenges for assessment implementation and provided recommendations towards the improvement of students' academic achievement.

Keywords: authentic assessment, traditional assessment, seafarer student, academic achievement, scores

1. Introduction

Past research (Maringa 2015; Emad and Roth 2007; AMC 2011) has demonstrated that seafarer (persons employed on ships) students tended to disengage with traditional assessments (e.g. multiple-choice questions, oral examinations, and written assignments) that focussed on their ability to recall and regurgitate the body of knowledge taught in the classroom. Disengaged students opted for surface-learning approaches (Maltby and Mackie 2009) relying on rote learning instead of assimilating and analysing information critically towards preparation for such assessment tasks. For example, one of the ways a seafarer is certified as competent to work onboard commercial ships is through an assessment based on memorised answers in an oral examination (Prasad 2011). Traditional assessments like oral examinations required students to construct responses devoid of context relying solely on the students' ability to visualise work-based scenarios. However, the ability to memorise is a lower-level cognition, and memory lapses may lead to unintentional skill and knowledge-based errors (Wiggins 1990) leading to poor academic achievement. Although traditional assessment methods may be effective in assessing knowledge-based components of a task, they are somewhat decontextualised in nature and find it difficult to provide students with a real-world context for skills and knowledge application (Boud and Falchikov 2006; Kearney 2012).

In the field of education, authentic assessment is presented as a model that requires students to provide responses to a situation described and delivered in a real-world (or contextually similar) contexts (Villarroel, Bloxham, Bruna et al. 2018). Authentic assessment tasks are found meaningful to students due to its strong figurative context and fidelity to the situations that they may find themselves in the professional world (Wiggins 1989; Gulikers 2006). Meaningful tasks set in real-world contexts enhance student engagement with assessment if students relate the tasks to professional practices (Richards-Perry 2011; Quartuch 2011). Past research (Brawley 2009; Schneider, Krajcik, Marx et al. 2001; Thomas 2000; Leon and Elias 1998; Gallagher, Stepien, and Rosenthal 1992) empirically proved higher student academic achievement for authentically assessed students when compared with their traditionally assessed counterparts. However, a literature review by Ghosh, Bowles, Ranmuthugala, and Brooks (2016; 2017) revealed that similar evidence is essentially missing in the area of seafarer education and training.

Hence, the objective of this research project was to investigate if authentic assessment implemented in seafarer education and training significantly increased students' academic achievement (through the comparison of scores obtained) as compared with traditional assessments. Separate and independent seafarer student groups were identified as the control (traditional assessment) and the treatment group (authentic assessment). The traditional and authentic assessments were implemented in the selected unit of 'Managerial and leadership skills' within the Bachelor of Nautical Science degree program at the Australian Maritime College (AMC), an institution of the University of Tasmania (UTAS). The Bachelor program of study is provided for students who intend to embark on a career in the maritime industry as ranked officers on commercial ships. It provides the knowledge and skills required to safely manage and operate ships. The unit of 'Managerial and leadership skills' was selected since it

enrolled the highest number of students within the degree program. Higher number of students maximized the participants and hence, enhanced the generalisability of findings.

The authentic assessment implemented for the selected unit differed from the decontextualised traditional assessment on the basis of the inclusion of a real-world context that attempted to closely replicate the complexities and challenges faced by seafarer students on ships. The inclusion of the real-world context being the only differing aspect between the two types of assessments, the 'authenticity' (provided through a real-world context) of the assessment was the focus variable.

However, according to past researchers (Bailey 1998, 205; Law and Eckes 1995; Dikli 2003, 16; Abeywickrama 2012) traditional assessments have been conventionally described as not only inauthentic but also as a "one-shot" and single-occasion tests implemented at the end of learning (summative) period. Hence, the scores obtained in the summative traditional assessments cannot inform on the progression of the learner as they only measure the students' ability at a particular time (Law and Eckes 1995). In comparison to the summative traditional assessments, one of the key characteristics of authentic assessment, as defined by its major authors (Wiggins 1989; Archbald 1991; Gulikers 2006), required students to be provided with more than one opportunity (formative) to apply their knowledge and skills. Students performing an authentic assessment task are provided with feedback on their performance and allowed to reflect on their work to recognize gaps in their knowledge. They are then, provided with another opportunity to perform a similar task at a different time (Law and Eckes 1995).

Hence, the authentic assessment in this research project was implemented as a formative assessment and in comparison, the traditional assessment was summative in nature. Since, the 'nature of task implementation' (formative versus summative) was a differing aspect between the two types of assessment, an additional variable (apart from 'authenticity') was introduced in this research. Hence, this research also investigated the difference in seafarer students' academic achievement comparing scores of the formative authentic assessment with the summative traditional assessment. Due to the nature of the assessment tasks (students were required to respond to questions based on a case study), additional independent variables (work experience, English as the first language, and educational qualification) based on their ability to influence student performance and the resulting academic achievement were also identified. The student scores were isolated on the basis of the independent variables and analysed to investigate the effect of these variables on students' academic achievement. The findings of this research project revealed recommendations for education and training providers towards the implementation of authentic assessment and improvement of students' academic achievement.

Research methodology

Research design

The difference in seafarer students' academic achievement (traditional versus authentic) for the unit of 'Managerial and Leadership Skills' was investigated in this research project. Students completing this unit acquire the knowledge and skills required by a senior seafarer officer to organise and manage the efficient operation onboard a merchant ship. The unit focuses on leadership and management of multicultural crews in a global environment and the maintenance of an effective interface with other industry stakeholders.

The students that enrolled in the unit in Semester 1 were classified as the 'control group' that underwent a traditional assessment. The traditional assessment comprised of two case study scenarios devoid of context. The students provided written responses to essay-type questions based on their analysis of the described scenarios presented devoid of context and relying on their ability to recall how the scenarios would have played out in the real-world onboard ships.

In comparison, another cohort of students enrolled in the same unit in Semester 2 were assessed authentically through the same case studies. Although, the authentically assessed students also provided written responses to the same essay-type questions, the authentic assessment differed from the traditional assessment by providing a real-world authentic context to the assessment task through a simulation and practical demonstration of the same case study scenarios, as employed in the traditional assessment, enacted by AMC staff. For example, one case study that described ship staff abandoning the ship using a liferaft during a fire was demonstrated at AMC training pool. The pool was equipped with facilities to launch a real liferaft in simulated waves, strong winds, darkness, rain, and smoke. The simulation also included ringing of the emergency alarms and staff playing the role of panicking seafarers jumping into the pool to replicate a possible emergency. In comparison to the authentic assessment, students assessed traditionally relied only on their imagination and experience to visualise the described scenarios.

In addition to the authentic design, the assessments also differed in the nature of their implementation. The authentic assessments were formative in nature and held on two different days (3 weeks apart). The second authentic task was implemented once the students received feedback on their performance in the first authentic task. In comparison to the authentic assessment, the traditional assessment was summative in nature and both case studies were implemented at the assessment. However, the duration of the authentic assessment (combined) was the same as that of the traditional assessment. The assessment details and rubric were provided to both the student groups at the beginning of the semester. To avoid the introduction of additional variables, the unit, learning content, lecture delivery methods, lecturer, assessment rubric, total duration of the assessment, and assessment questions were kept constant. Both the assessments were supervised by external invigilators appointed by AMC. Table 1 summarises the research design.

Table 1: Summary of research design.

Unit of competence	Managerial and leadership skills	Managerial and leadership skills
Participants	Seafarer students enrolled in the Bachelor of Nautical Science degree program	Seafarer students enrolled in the Bachelor of Nautical Science degree program
Group	Control Group	Treatment group
Semester	1	2
Assessment type	Traditional assessment	Authentic assessment
Task description	Students respond to case studies described in the assessment	Students are provided with a real-world 'authentic' context for the case study described in the assessment
Nature of assessment	Summative	Formative
Task implementation	Two case studies implemented together	One case study implemented three weeks apart (Total: 2 case studies)
Response method	Written response to essay-type questions	Written response to essay-type questions
Duration	One hour	30 minutes for each case study (Total: 1 hour)

Data analysis

The quantitative data (assessment scores) was analysed using MS Excel. The student scores were analysed using the values of mean scores, standard deviation, effect size, and the t-test values. While the mean scores provided an indication on the difference in students' academic achievement between the two types of assessments implemented, standard deviation informed on the scattering of the individual scores in each type of assessment to indicate the variation. The recommended (Coe 2002) effect size (0.5 or greater) and the t-test values ($P < .05$) indicated if the variation in scores of students' academic achievement was statistically significant for reporting.

Sampling considerations

The sampling technique used in this research was based on convenience sampling that relies on opportunity and participant accessibility, and used when the study population is large, and the research is unable to test every individual (Clark 2014; Robson 2011). Participants for this research were seafarer students drawn from the course of Bachelor of Nautical Science enrolled in the selected unit at AMC. This research was based on the sample of 96 participants (as the control group), and 93 participants (as the treatment group). Scores of seven students from the control group and nine students from the treatment group were not included in the analysis due to the failure of the students to complete the administrative paper work. A key consideration while sampling was to ensure that the control and treatment groups comprised of randomly assigned students where each participant had an equal chance of participating in this research based only on the sequence of enrolment in the individual semesters. The groups

were not sorted based on any other pre-determined characteristics, such as qualifications, academic ability or work experience that may have impacted the outcomes of this research.

Validity and reliability of assessment

Before implementing assessment

Content and construct validity was achieved by using a jury of experts before the assessment was implemented. The subject experts comprised of seven field experts within AMC. The subject experts included ex-seafarers currently employed as educators in the field of seafarer education, each having more than 25 years of work experience. The first draft of the assessment instrument was sent to the subject experts who were asked to make recommendations towards improving the instruments. The experts provided suggestions on simplifying terms used in the case study for universal understanding. For example, the words 'imperative', 'mitigate', and 'hinder' were substituted with the words 'vital importance', 'reduce', and 'delay'. Suggestions were also provided on the distribution of marks, length of the tasks, and ways to demonstrate the case studies authentically within the educational settings at AMC.

After implementing assessment

Criterion validity for authentic assessment was obtained with a secondary authentic assessment implemented three weeks after the first assessment. The test for criterion validity allowed to assess the consistency of student performance in authentic assessments.

To establish more consistency, objectivity, and reliability, the student scores were reviewed by the panel of the subject experts using the assessment rubric.

Ethics approval

A minimal risk ethics application approval, constituting ethical clearance by the Tasmania Social Sciences Human Research Ethics Committee, was obtained for this research project. Participants were reassured that the data would be anonymised and that their contribution would be confidential. Students were free to withdraw at any time from the study.

Research questions, hypothesis, and independent variables

The focus of this research project was to investigate the difference in seafarer students' academic achievement by comparing traditional assessment scores with authentic assessment scores. Hence the following research questions (RQ) were developed:

RQ1: Is there a significant improvement in seafarer students' academic achievement in authentic assessment when its scores are compared with traditional assessment scores?

RQ1 enabled the development of the following research hypothesis (H_{1a}):

1a) H_{1a}: There is a significant improvement in seafarer students' academic achievement in authentic assessment when its scores (AA) are compared with traditional assessment scores (TA). This hypothesis was denoted as: *Score AA > Score TA*

As stated in the 'Research design' section, the authentic assessment was implemented as two separate tasks (or case studies) with the second task implemented three weeks after the first task. The authentically assessed students received feedback on their performance in the first task before attempting the second authentic task three weeks later. In comparison, the traditional assessment implemented both tasks at the same assessment. Hence, the traditionally assessed students did not receive feedback on the first task to improve their performance in the second task. Since, the first task in both traditional and authentic assessments were performed without any prior feedback, and the differing aspect between the assessments was only the 'authentic' nature, the next hypothesis was also developed towards answering RQ1.

1b) H_{1b}: There is a significant improvement in seafarer students' academic achievement in authentic assessment when its scores for the first task (AA_1) are compared with traditional assessment scores for the first task (TA_1). This hypothesis was denoted as:
Score AA_1 > *Score* TA_1

It was evident from the differing nature of assessment implementation (formative versus summative) that contrary to students assessed authentically, students assessed traditionally did not receive more than one opportunity to improve their academic achievement based on feedback. Thus, apart from the 'authentic' design, additional variables (e.g. multiple opportunity based on feedback) that may have influenced student achievement in this research were introduced due to the nature of assessment implementation. Hence, the difference in seafarer student achievement by comparing scores obtained in summative traditional assessment with scores obtained in formative authentic assessment was investigated in this research project. This resulted in the development of the following RQ:

RQ2: Is there a significant improvement in seafarer students' academic achievement in the formative authentic assessment when its scores are compared with summative traditional assessment scores?

RQ2 enabled the development of the following research hypothesis:

2a) H_{2a}: There is a significant improvement in seafarer students' academic achievement in authentic assessment when its scores for the second task (AA_2) are compared with traditional assessment scores for the second task (TA_2). This hypothesis was denoted as: *Score* AA_2 > *Score* TA_2

To answer RQ2, it was necessary to investigate the difference in the students' academic achievement if the assessment design was kept constant, and the only differing aspect between the student performances was the nature of assessment implementation. It was assumed that authentically assessed students that received more than one opportunity and feedback on their performance in the first task, would achieve higher scores in the second task. Hence, keeping the 'authentic' design of the assessment as a constant, the following hypothesis was developed:

2b) H_{2b}: There is a significant improvement in seafarer students' academic achievement in authentic assessment when its scores for the second task (AA_2) are compared with the scores for the first task (AA_1). This hypothesis was denoted as: $Score\ AA_2 > Score\ AA_1$

Since, the summative nature of the traditional assessment did not allow students to receive feedback on their performance in the first task to recognize gaps in their knowledge; and another opportunity to improve their academic achievement in the second task, it was assumed that traditionally assessed students would find it challenging to significantly improve their academic achievement in the second task. Hence, keeping the 'traditional' design of the assessment as a constant, the following hypothesis was developed:

2c) H_{2c}: There is no significant improvement in seafarer students' academic achievement in the second traditional assessment task (TA_2) when its scores are compared with the scores for the first task (TA_1). This hypothesis was denoted as: $Score\ TA_2 \sim Score\ TA_1$

The research questions and the resulting hypotheses is summarised in Table 2.

Table 2: Research questions and the resulting hypothesis.

Research Question	Hypothesis
RQ1: Is there a significant improvement in seafarer students' academic achievement in authentic assessment when its scores are compared with traditional assessment scores?	H_{1a}: $Score\ AA > Score\ TA$ H_{1b}: $Score\ AA_1 > Score\ TA_1$
RQ2: Is there a significant improvement in seafarer students' academic achievement in formative authentic assessment when its scores are compared with summative traditional assessment scores?	H_{2a}: $Score\ AA_2 > Score\ AA_1$ H_{2b}: $Score\ AA_2 > Score\ TA_2$ H_{2c}: $Score\ TA_2 \sim Score\ TA_1$

Independent variables

This research focussed on investigating the difference in students' academic achievement that may have resulted either due to the design of the assessment (traditional versus authentic) or the nature of its implementation (summative versus formative). Hence, this research required to identify the independent variables that could influence the achievement. Since, both traditional and authentic assessments implemented in this research required seafarer students to respond to case study scenarios, the independent variables identified were based on their efficacy with regards to influencing student performance and resulting scores. Thus, the following variables were identified:

- **Work experience**: The assessment tasks required students to respond to case study scenarios based on situations that they might encounter on board ships. There was a possibility that students with higher work experience may have encountered similar situations and hence, were better equipped to answer the questions. Although it was not a stringent requirement, students enrolled in the selected unit were expected to

have completed the minimum work experience of one and half to three years on ships. Thus, the extraneous variable of 'work experience' was classified as students with 'less than three years' and 'more than three years' of experience.

- **English as first language**: Since students were required to provide written responses describing their actions in the case study scenarios, proficiency in the English language could significantly affect their ability to provide descriptive answers. This research project does not imply that all non-native English speakers do not have proficiency over the language. Since, this project did not conduct any additional tests to assess the English language proficiency of non-native English speakers, it was necessary to distinguish them from students with English as their first language.
- **Level of education completed**: The minimum requirement for enrolment in the bachelor's program is a senior secondary school (Grade 10 - Grade 12) qualification. However, the selected sample for this research included students with qualifications higher than Grade 12 including those with under-graduate or post-graduate qualification from universities. Students completing higher academic qualifications such as university studies may be better equipped in their ability to analyse and respond to case study scenarios compared with students who have only completed studies at school level. Hence, the variable of 'level of education completed' was classified as students who had completed up to high school (Grade 10- 12) and students who had completed education higher than Grade 12.

Results

The results were analysed against the RQs and corresponding hypothesis described previously. Findings are summarised in Table 3.

Table 3: Summary of research findings for RQ1 and RQ2.

RQ	Hypothesis	Findings
RQ1	H_{1a}: Score AA > Score TA	Student achievement was significantly higher in authentic assessment for the composite group and groups isolated on independent variables.
RQ1	H_{1b}: Score AA_1 > Score TA_1	Student achievement was significantly higher in authentic assessment for the composite group and groups isolated on independent variables (except for student groups with more than three years of work experience).
RQ2	H_{2a}: Score AA_2 > Score TA_2	Student achievement was significantly higher in authentic assessment for the composite group and groups isolated on independent variables.
RQ2	H_{2b}: Score AA_2 > Score AA_1	Student achievement was significantly higher in the second authentic task for the composite group and groups isolated on independent variables.
RQ2	H_{2c}: Score TA_2 ~ Score TA_1	No significant difference in seafarer student achievement found for the composite group and groups isolated on independent variables.

The results are presented in the following section.

RQ1: Is there a significant improvement in seafarer students' academic achievement in authentic assessment when its scores are compared with traditional assessment scores?

Table 4 provides a result summary for RQ1.

Table 4: Result summary for RQ1.

Hypothesis	Mean Score	Sample	S.D.	Effect Size	t-Test (df); P (two-tail) assuming unequal variance
1a) Score AA > TA	AA (69.8%) > TA (52.5%)	AA (93) TA (96)	AA (14.6) TA (20.6)	0.98	t (172) 6.7; P <.05
1b) Score AA_1 > TA_1	AA_1 (63.8%) > TA_1 (52.4%)	AA_1 (93) TA_1 (96)	AA_1 (9.6) TA_1 (11.2)	0.55	t (184) 3.8; P <.05

Table 4 showed that AA significantly improved by 17.3% compared with TA; and AA_1 was 11.4% higher than TA_1. The hypotheses (H_{1a} and H_{1b}) designed for RQ1, thus, held true. In both hypotheses, the S.D. values indicated higher scattering amongst traditional assessment scores; and the effect size and the t-test values showed that the difference and variation in the scores were significant for reporting.

RQ2: Is there a significant improvement in seafarer students' academic achievement in formative authentic assessment when its scores are compared to summative traditional assessment scores?

Table 5 provides a result summary for RQ2.

Table 5: Result summary for RQ2.

Hypothesis	Mean Score	Sample	S.D.	Effect Size	t-Test (df); P (two-tail) assuming unequal variance except *
2a) Score AA_2 >TA_2	AA_2 (75.8%) > TA_2 (52.6%)	AA_2 (93) TA_2 (96)	AA_2 (7.3) TA_2 (10.7)	1.2	t (168) 8.7; P <.05
2b) Score AA_2 >AA_1	AA_2 (75.8%) > AA_1 (63.8%)	AA_2 (93) AA_1 (93)	AA_2 (7.2) AA_1 (9.6)	0.71	t (171) 4.8; P <.05
2c) Score TA_2 ~ TA_1	TA_2 (52.4%) ~ TA_1 (52.6%)	TA_2 (96) TA_1 (96)	TA_2 (11.2) TA_1 (10.7)	0.01	t (190) .11; P > .05 *assuming equal variance

Analysis of the composite scores as presented in Table 5 showed the following:

- AA_2 significantly improved by 23.2% when compared with TA_2;
- AA_2 significantly improved by 12% when compared with AA_1; and
- no significant difference was found between TA_1 and TA_2.

The hypotheses (H_{2a}, H_{2b}, and H_{2c}) designed for RQ2, thus, held true. In hypothesis H_{2a} and H_{2b}, the S.D. values indicated that the TA_2 and AA_1 scores were more widely scattered than the AA_2 scores; and the effect size and the t-test values showed that the difference and variation in scores was significant for reporting. In hypothesis H_{2c}, due to the similarity in the composite mean score values of TA_1 and TA_2, as expected, the S.D. values indicated that the scores of both the traditional tasks were similarly scattered; and the effect size and the t-test values showed that the difference and variation in scores were not significant for reporting.

The effect of independent variables on students' academic achievement

H_{1a}: Score AA > Score TA

Table 6 summarises the effect of the independent variables on H_{1a}.

Table 6: Effect of independent variables on H_{1a}.

H_{1a}: Score AA > TA	**Mean Score AA > TA (%)**	**Sample AA/TA**	**S.D. AA/TA**	**Effect Size**	**t-Test (df); P (two-tail) assuming unequal variance**
Composite:	69.8 > 52.5	93/96	14.6/20.6	0.98	t (172) 6.8; P <.05
Work Experience:					
< 3 years	65.5 > 44.8	31/61	18.1/19.4	1.10	t (64) 5.0; P < .05
> 3 years	72.0 > 65.7	62/35	12.1/15.4	0.50	t (58) 2.1; P < .05
English (first language):					
Yes	79.3 > 64.0	33/27	9.8/15.3	1.20	t (43) 4.5; P < .05
No	64.7 > 48.0	60/69	14.3/20.7	0.95	t (121) 5.4; P < .05
Educational qualifications:					
High school	69.4 > 49.8	45/68	15.1/19.3	1.10	t (108) 6.0; P < .05
University	70.3 > 58.9	48/28	14.3/22.5	0.62	t (40) 2.4; P < .05

Table 6 showed that the scores, when isolated on the independent variables, revealed that the students' academic achievement significantly improved in AA when compared with TA. The hypothesis (H_{1a}), thus, held true for all the independent variables. The S.D. values indicated that the TA scores were more widely scattered than the AA scores; and the effect size and the t-test values showed that the difference and variation in scores was significant for reporting for all the independent variables.

H_{1b}: Score AA_1 > Score TA_1

Table 7 summarises the effect of the independent variables on H_{1b}.

Table 7: Effect of independent variables on H_{1b}.

H_{1b}: Score AA_1 > TA_1	**Mean Score AA_1 > TA_1 (%)**	**Sample AA_1/TA_1**	**S.D. AA_1/TA_1**	**Effect Size**	**t-Test (df); P (two-tail) assuming unequal variance**
Composite:	63.8 > 52.4	93/96	9.6/11.2	0.55	t (184) 3.8; P < .05
Work Experience:					
< 3 years	58.6 > 44	31/61	10.5/11.3	0.67	t (57) 3.0; P < .05
> 3 years	66.6 ~ 66.6	62/35	8.5/8.6	0.00	t (70) 0.0; P > .05
English (first language):					
Yes	73.6 > 64.2	33/27	7.9/9.4	0.54	t (51) 2.0; P < .05
No	58 > 47.6	60/69	9.4/11.1	0.51	t (127) 2.9; P < .05
Educational qualifications:					
High school	62 > 49.8	45/68	10.3/11.0	0.74	t (99) 3.0; P < .05
University	65.6 > 58.2	48/28	9.0/11.3	0.63	t (47) 1.5; P < .05

Table 7 showed that the scores, when isolated on the independent variables, revealed that the students' academic achievement significantly improved in AA_1 when compared with TA_1. The only exception was in the case of students with more than three years of work experience where the scores were found to be similar in value. This indicated that for the first task, traditionally assessed students with more than three years of work experience benefitted from their familiarity with the workplace, related the assessment task to the real-world context and hence, were able to respond as well as the authentically assessed students.

The hypothesis(H_{1b}), thus, held true for all the independent variables but with a single exception. The S.D values of TA_1 were more widely scattered than the AA_1 scores; and the effect size and the t-test values showed that the difference and variation in scores were significant for reporting in all groups isolated on the independent variables except for students with more than three years of work experience. Due to similarity in the AA_1 and TA_1 scores of students with more than three years of work experience, the S.D. values indicated similar scattering; and the effect size and the t-test values showed that the difference and variation in scores were insignificant for reporting.

H_{2a}: Score AA_2 > Score TA_2

Table 8 summarises the effect of the independent variables on H_{2a}.

Table 8: Effect of independent variables on H_{2a}.

H_{2a}: Score AA_2 > TA_2	**Mean Score AA_2 > TA_2 (%)**	**Sample AA_2/TA_2**	**S.D. AA_2/TA_2**	**Effect Size**	**t-Test (df); P (two-tail) assuming unequal variance**
Composite:	75.8 > 52.6	93/96	7.3/10.7	1.20	t (168) 8.7; P < .05
Work Experience:					
< 3 years	72.6 > 45.6	31/61	8.8/10.0	1.40	t (68) 6.6; P < .05
> 3 years	77.6 > 64.8	62/35	6.2/8.9	0.84	t (53) 3.7; P < .05
English (first language):					
Yes	82.8 > 63.8	33/27	5.2/8.8	1.30	t (41) 4.9; P < .05
No	71.4 > 48.2	60/69	7.4/10.6	1.50	t (122) 7.2; P < .05
Educational qualifications:					
High school	76.8 > 49.8	45/68	6.4/9.8	1.70	t (111) 8.8; P < .05
University	75 > 59.6	48/28	8.0/12.0	0.76	t (41) 3.0; P < .05

The scores, when isolated on the independent variables, revealed that the students' academic achievement significantly improved in AA_2 when compared with TA_2. The hypothesis(H_{2a}), thus, held true for all the independent variables. The S.D. values indicated that the TA_2 scores were more widely scattered than the AA_2 scores; and the effect size and the t-test values showed that the difference and variation in scores was significant for reporting for all the independent variables.

H_{2b}: Score AA_2 > Score AA_1

Table 9 summarises the effect of the independent variables on H_{2b}.

Table 9: Research findings to H_{2b}.

H_1: Score AA_2 > AA_1	**Mean Score AA_2 > AA_1 (%)**	**Sample AA**	**S.D. AA_2/AA_1**	**Effect Size**	**t-Test (df); P (two-tail) assuming unequal variance**
Composite:	75.8 > 63.8	93	7.2/9.6	0.71	t (171) 4.8; P < .05
Work Experience:					
< 3 years	72.6 > 58.6	31	8.8/11.3	0.69	t (57) 2.7; P < .05
> 3 years	77.6 > 66.6	62	6.2/8.5	0.74	t (113) 4.1; P < .05
English (first language):					
Yes	82.8 > 73.6	33	5.2/7.9	0.70	t (56) 2.8; P < .05
No	71.4 > 58	60	7.4/9.4	0.79	t (112) 4.3; P < .05
Educational qualifications:					
High school	76.8 > 62	45	6.4/10.3	0.88	t (74) 4.1; P < .05
University	75 > 65.6	48	8.0/9.0	0.55	t (93) 2.7; P < .05

Table 9 showed that the scores, when isolated on independent variables, revealed that the students' academic achievement significantly improved in AA_2 when compared with AA_1. The hypothesis (H_{2b}), thus, held true for all the independent variables. The S.D. values indicated that the AA_1 scores were more widely scattered than the AA_2 scores; and the effect size and the t-test values showed that the difference and variation in scores was significant for reporting.

H_{2c}: Score TA_2 ~ Score TA_1

Table 10 summarises the effect of the independent variables on H_{2c}.

Table 10: Research findings to H_{2c}.

H_1: Score TA_2 ~ TA_1	Mean Score TA_2 ~ TA_1	Sample TA	S.D. TA_2/TA_1	Effect Size	t-Test (df); P (two-tail) assuming equal variance
Composite:	52.6 ~ 52.4	96	10.7/11.2	.01	t (190) .11; P > .05
Work Experience:					
< 3 years	45.6 ~ 44	61	10.0/10.5	.08	t (120) .42; P > .05
> 3 years	64.8 ~ 66.6	35	8.9/8.6	.10	t (68) .42; P > .05
English (first language):					
Yes	63.8 ~ 64.2	27	8.8/9.4	.03	t (52) .09; P > .05
No	48.2 ~ 47.6	69	10.6/11.0	.02	t (136) .18; P > .05
Educational qualifications:					
High school	49.8 ~ 49.8	68	9.8/11.0	.00	t (134) .01; P > .05
University	59.6 ~ 58.2	28	12.0/11.3	.06	t (54) .22; P > .05

Table 10 showed that the scores, when isolated on the independent variables, revealed that there was no significant difference in students' academic achievement when TA_1 was compared with TA_2. The hypothesis (H_1: Score TA_2 ~ Score TA_1), thus, held true for all the independent variables. Although, the TA_2 score values was not always exactly equal to the TA_1 score values, the maximum difference between the two scores did not exceed 2% which was not considered significant in this research project. Due to the similarity in the mean score values of TA_1 and TA_2, as expected, the S.D. values indicated that the scores of both the traditional tasks were similarly scattered; and the recommended (Coe 2002) effect size and the t-test values showed that the difference and variation in scores was insignificant for reporting.

Discussion

Higher academic achievement in authentic assessment

The results of this research project confirmed that seafarer students' academic achievement in authentic assessment is higher when compared with traditional assessment. This finding indicated that the academic achievement of seafarer students improved significantly when their responses to the questions in the assessment task were not relying on memorisation of information and imaginings of a situation but on the assimilation, integration, and analysis of information provided in a real-world context. The real-world context in which the authentic assessment tasks were based in this project mirrored a task that was situated and required to be solved 'similar' to what is faced in professional life. The response to the questions in the tasks, then, reflected work or demonstrated knowledge and skills as would be required at the

workplace. This enabled seafarer students to relate classroom learning to the work onboard ships, leading to higher academic achievement.

Higher academic achievement by seafarer students in this project corroborated the findings of non-seafarer research (Brawley 2009; Schneider, Krajcik, Marx et al. 2001; Thomas 2000; Leon and Elias 1998; Gallagher, Stepien, and Rosenthal 1992) where students assessed authentically demonstrated higher achievement in comparison to traditionally assessed students. Although, the findings were similar, this research made a unique contribution by studying participants in post-school settings compared with past research that was conducted in the educational settings of a school, e.g. Brawley (2009) (early childhood defined as grade pre-K to third); Schneider et al. (2001) (tenth and eleventh grade); Thomas (2000) (tenth grade); Leon and Elias (1998) (sixth grade); and Gallagher et al. (1992) (unspecified school students).

This research also distinguished itself from past research by using two separate student groups as the 'control' and 'treatment' group. Past research (e.g. Brawley 2009) used the same student group for both traditional and authentic assessments. In cases where the same group of students are used for both assessment, the higher achievement of the students transitioning from traditional to authentic assessment may be attributed to the 'learning effect' This refers to the gain in student knowledge that may have occurred in the time between the administrations of the two assessments. Learning effect creates an additional variable, which was avoided in the research designed for this project.

Although, past researchers like Schneider et al. (2001); Thomas (2000); Leon and Elias (1998); and Gallagher et al. (1992) used two separate randomly assigned groups for comparison between authentic and traditional assessment performance, additional variables other than the 'authentic' design of the assessment may have been introduced due to the nature of the tasks or associated learning. For example, Leon and Elias (1998) used 'portfolios' versus 'self-selected performance-based projects; Gallagher et al. (1992) used 'open-ended questions' versus 'authentic performance task'; and Schneider et al. (2001) and Thomas (2000) used two separate groups with a different learning experience before the authentic and traditional tasks were administered. In contrast, the focus of this research was only on studying the impact of the 'authentic' design and the 'nature of implementation' (as the differing aspects between the two types of assessment) on students' academic achievement keeping the remainder of the variables (e.g. learning content, assessment questions and duration, etc.) constant.

Higher academic achievement in formative assessment

The findings of this project also confirmed that seafarer students' academic achievement improved in formative assessment when compared with summative assessment. The formative assessment employed in this research project provided students with an opportunity to receive feedback on their performance in the first authentic assessment task before attempting the second authentic assessment task. According to Zhang and Zheng (2018), feedback on a students' current ability to perform an assessment task and providing suggestions to improve and attain expected levels, encourages students to take necessary actions to close the gap in

their ability. This was confirmed empirically in this project. For example, higher academic achievement in AA_2 as compared with AA_1 indicated that using the feedback obtained, seafarer students recognized the gaps in their knowledge, re-evaluated their learning approaches and implemented new strategies to improve their scores. In comparison to the formative assessment, the feedback obtained by the students in the summative traditional assessment task proved to be too late for the control group of seafarer students to make any adjustments to their learning process to improve their scores.

In ascertaining the major influences on student achievement, Hattie (2009) synthesised more than 800 meta-analyses in education and concurred that one of the key requirements for skills improvement is feedback on the students' current level of skills and multiple opportunities to practice the skills. This was also empirically reconfirmed in this project and evident in the scores of seafarer students in the comparison of AA_1 to TA_1 and AA_2 to TA_2, when isolated for the independent variable of 'more than three years of work experience'. The scores of AA_1 were similar to TA_1 when isolated for the specified variable which indicated that seafarer students with higher work experience may negate the advantage provided through the real-world context of authentic assessments due to their experience in performing similar tasks in the workplace. However, the scores of AA_2 were significantly higher than TA_2 when isolated for the same variable. This suggested that the factor of 'higher work experience' could not nullify the advantage provided through the real-world context in the second authentic assessment task. Since, the comparison was between the same group of students for both tasks, the only advantage provided to the authentically assessed students over the traditionally assessed students for the second assessment task was a feedback and an opportunity to improve on their performance and resulting scores.

Higher academic achievement in a formative authentic assessment as compared with a summative traditional assessment, as found in this project, corroborated the past research findings of Johnson (2007). Similar to this project, Johnson (2007) used two randomly assigned student groups (traditional assessment as control and authentic assessment as treatment); and measured student achievement using assessment scores. However, the research findings of Johnson (2007) was based on school students. Moreover, Johnson's (2007) research compared traditional and summative multiple-choice assessments with a formative authentic assessment. Multiple-choice assessments have a different format (solution to problem provided as a choice of selection) to performance-based authentic assessments (solution needs to be constructed from information provided). In such cases where the formatting of the assessment questions may influence student performance, additional variables other than assessment design (correct solution may be a result of guesswork and not an understanding of the problem) may be introduced to which the higher student achievement may be attributed to. Additional to assessment format, in Johnson (2007), students in the control group were also asked to complete the assessment by hand using whereas the treatment group was tested using authentic assessments via the computer. The differing modes of answering, for example, the role of technology in task performance as acknowledged by Johnson (2007), may have added a variable that impacted student achievement. In comparison, this research ensured that the student performance is only affected by the focus variables (authenticity of the assessment and/or nature of assessment implementation) and not by any other unexplained variables.

Impact of independent variables on academic achievement in authentic assessment

In this research project, the analysis of the student scores within the control and treatment groups (overall and when isolated on independent variables) revealed that student achievement was significantly higher for students with more than three years of work experience as compared with students with less than three years of work experience. This implied that students with lesser work experience may not have had adequate experience at their workplace performing tasks similar to the assessment tasks designed for this project. In comparison, there is a possibility that students with more than three years of work experience were more familiar with the assessment tasks and had performed them in the workplace contexts which enabled them to significantly score higher than their less experienced counterparts. Educators face the challenge of teaching and assessing students with differing work experience within the same cohort. In such cases, educators should strive for parity in student ability to perform the task via greater opportunities to practice similar tasks before the main assessment. Teacher feedback on practice attempts will allow students to identify their areas of weakness and address them. Due to time constraints, one of the limitations of this project was its' inability to provide students with an opportunity to practice similar tasks before the main assessment.

Analysis of the student scores also revealed that student achievement was significantly higher for students with English as their first language as compared with their non-native English-speaking counterparts within both control and treatment groups (overall and when isolated on independent variables). One of the key reasons for this finding may be attributed to the format of the assessment which required students to respond to questions based on a case-study. Answering the questions in English may have affected the performance of the students who were not proficient in the language, and hence, lowered their academic achievement. In countries where training and assessments are conducted using the English language (for example, this project was set in Australia), a key implication for educators lie in facing the challenge of teaching and assessing students with differing abilities in communicating using the English language. Educators must investigate ways to design authentic assessments that require students to perform tasks that require more hands-on approach with lesser focus on language abilities. Another solution could also be to involve students in the design of the authentic assessment. Including student voice will address respective concerns and allow educators to plan for them in advance.

Analysis of the scores (overall and when isolated on independent variables) within the treatment group revealed that the educational qualifications of a student had no significant impact on student achievement in authentic assessment. For example, in both the first and second authentic assessment tasks, students with university level of education did not score significantly higher (the difference of marks was less than two percent) than the students with only high school qualifications. This finding indicated that authentic assessment enacted in real-world settings evened out student ability in analysing contexts for critical assimilation of information towards providing response to assessment questions.

In comparison analysis of the scores within the control group (overall and when isolated on independent variables), revealed that students with university level of qualification scored significantly higher than students with only high school qualifications. In absence of a real-world contexts, students in traditional assessment relied on their ability to analyse

information based on their ability to read and comprehend a descriptive case-study. Hence, students with university qualifications used their academic experience of participating in similar context-devoid assessments and scored over their lesser qualified (high school) counterparts. This finding might be considered to contradict the findings of Johnson (2007) who based on a similar analysis found that authentically assessed students on the gifted and honours track scored significantly higher than students on the technical (vocational) track. This obviously assumes that we can equate 'gifted and honours track' with the independent variable of 'University education'; and 'technical track' with 'high school'. Although, it is likely that the relationship between educational attainment and academic scores in assessment is less than perfect, educators must investigate ways to design authentic assessment to bridge the important gap between students with differing educational backgrounds.

Contribution in the field of seafarer education and training

Extensive literature reviews conducted in past research (Ghosh et el. 2016; 2017) revealed that there is a global absence of empirical evidence on authentic assessment in seafarer education. This research investigated the difference in seafarer students' academic achievement; and found higher academic achievement in authentic assessment when its scores were compared with traditional assessment scores. In doing so, much- needed empirical evidence on the impact of authentic assessment in seafarer education was collected.

Disadvantages of implementing authentic assessment

Although the implementation of authentic assessment in this project provided a significant advantage through higher academic achievement for seafarer students, the discussion section on this project would not be completed without the inclusion of the analysis conducted on the disadvantages of implementing authentic assessment. Past research (Neely and Tucker 2012; Wiggins 1989) suggested that the direct and indirect costs of developing authentic assessments is twice as much as traditional assessments. Day, Blankenstein, Westenberg et al. (2018) claimed that authentic assessments are often time consuming. Hence, a comparative cost and time analysis for the resources used in developing the traditional and authentic assessment was conducted for this research. Table 11 details the differences in the resources used and the costs incurred.

Table 11: Comparison of cost estimation in assessment implementation.

SNo.	Resources	Traditional Assessment Costs	Authentic Assessment Costs
1.	Staff	1 lecturer = \$180/hour 2 invigilators x \$40 = \$80/hour **Total cost (TA_1 and TA_2) = \$260**	1 lecturer = \$180/hour 4 staff members to demonstrate authentic case studies x \$180 = \$720 2 invigilators x \$40 = \$80/hour Cost for AA_1 = \$980 Cost for AA_2 = \$980 **Total cost (AA_1 and AA_2) = 2 x \$980 = \$1960**
2.	Classrooms	1 classroom (TA_1 and TA_2) x \$100/hour = \$100 **Total cost (TA_1 and TA_2) = \$100**	1 classroom (AA_1) x \$100 = \$100 1 classroom (AA_2) x \$100 = \$100 **Total cost (AA_1 and AA_2) = \$200**
3.	Facilities	**None**	AMC Pool; liferaft; smoke generators; safety equipment = \$600/day Lifeguard = \$40/hour Cost for AA_1 = \$640 Cost for AA_2 = \$640 **Total cost (AA_1 and AA_2) = 2 x \$640 = \$1280**
4.	Time	1 hour for implementing both tasks; Both tasks completed in one day; 2 hours for providing students with common feedback on their performance; 2 hours of assessor's meeting **Total time used = 5 hours**	2 hours for developing and demonstrating both authentic case studies; 2 hours for implementing both tasks; Both tasks completed in three weeks; 4 hours for providing students with common feedback on their performance; 4 hours of assessor's meeting **Total time used = 12 hours**
5.	Answer booklets	100 booklets (TA_1 and TA_2) x \$0.70 = \$70 **Total cost (TA_1 and TA_2) = \$70**	100 booklets (AA_1) x \$0.70 = \$70 100 booklets (AA_2) x \$0.70 = \$70 **Total cost (AA_1 and AA_2) = \$140**

In Table 11, the time and costs (shown in Australian dollars) analysis conducted for this project showed that compared with traditional assessment, authentic assessment employed resources at a significantly higher cost. This confirmed the theoretical claims of Neely and Tucker (2012), Wiggins (1989), and Day et al. (2018). In this project, the cost of implementing a new and innovative assessment (authentic assessment) was significantly more than maintaining an existing assessment (traditional assessment). For such cases, Joughin, Dawson, and Boud (2017) argued that changes in the assessment regime become justified if the benefits outweigh the costs. This project provides the necessary evidence through higher academic achievement for authentically assessed seafarer students, and in doing so justifies its implementation in education and training.

Conclusion

A rigorous experimental design to conduct a comparative study of seafarer students' academic achievement between traditional and authentic assessment was set up in this research project. The research was designed by isolating the 'authentic' element in assessment to study its impact on scores obtained. Additional to the 'authentic' element, this research was also designed to conduct a comparative analysis of students' academic achievement between a formative (authentic) and a summative (traditional) assessment. To ensure that the research

outcomes were not influenced by any unidentified bias, the independent variables that could possibly affect student performance was also identified. The impact of the identified variables was studied separately to accurately measure their impact.

This research made its contribution through the collection of much needed empirical evidence on the impact of authentic assessment in seafarer education since similar research has not been conducted before globally. On the basis of the comparison between authentic and traditional assessment scores, the findings of this research confirmed that seafarer students assessed authentically had significantly higher scores resulting in higher academic achievement. This finding indicated that students' academic achievement will be improved if they focus on the assimilation, critical analysis, and integration of information presented through a real-world context instead of memorising information and rote learning.

The findings of this research also confirmed higher academic achievement for seafarer students in a formative (authentic) assessment when compared with a summative (traditional) assessment. This finding indicated that students' academic achievement will be improved if they are provided with feedback that may be used in recognizing gaps in their knowledge and skills and then at least one opportunity to attempt a similar task before the judgement on their competence is made. In the context of seafarer education, a shift is hence required from summative oral assessments that declare students as 'fail' before being provided with a feedback or another opportunity.

The use of summative examinations at the end of the learning period represents the final judgement of the students' performance and is often too late to make any changes to the learning strategies. However, educators should provide timely and efficient feedback to students; and receive counter feedback to reflect on areas they may improve upon as well. Students should take advantage of the feedback and work closely with the educators and assessors to become active participants in the learning process by recognizing their strength and weaknesses; and in establishing realistic learning goals. This develops their metacognitive ability of reflecting on their current learning practices and improving on them. Reflection on practices is a critical part of professional performance required to avoid errors.

Authentic assessment implemented in this research required the assistance of additional staff members employed at AMC. This suggested that embedding authentic assessment in a course may require cross-disciplinary teams to work closely together. This differs from the current work allocation methods of one lecturer per subject. Additionally, to educators working together, policies of education institutes need to support the organization of funds and other required resources towards assessment implementation.

This research highlighted that educators face the challenge of assessing students with different work experiences, proficiency in the English language, and educational qualifications. The authentic assessment employed in this project was able to achieve parity in performance and resulting scores only between students with university and high school qualifications. In all other cases of authentic assessment, and for the traditionally assessed students, academic achievement was higher for students with higher work experience, proficiency in the English language, and educational qualifications. To address the needs of the learners with different backgrounds and achieve equity in academic achievement, this project recommends educators to provide students with the opportunity to practice tasks

similar to the assessment tasks. Inability to do so due to time constraints, is one of the key limitations of this research.

This research collected data from the educational programme conducted at AMC due to the convenience in the selection and recruitment of participants. In future, this research should be replicated in other seafarer training institutes and for other units of competence to enhance generalisability of findings. In this research, higher academic achievement was attributed to the 'authentic' design of the assessment; and the formative nature of its implementation. However, further research is required to investigate other factors of assessment (for example, transparency of assessment criteria, assessment task, feedback, etc.) that seafarer students may have perceived significant towards their higher achievement. To do so, future research will correlate seafarer students' perception of authenticity in assessment to their scores in the assessment task. The factors correlating significantly will be included in designing assessment tasks towards improving the academic achievement.

7.7. Paper VII

INVESTIGATING THE CORRELATION BETWEEN STUDENTS' PERCEPTION OF AUTHENTICITY IN ASSESSMENT AND THEIR ACADEMIC ACHIEVEMENT IN THE ASSOCIATED ASSESSMENT TASKS

Samrat Ghosh
Australian Maritime College, University of Tasmania
Benjamin Brooks
Australian Maritime College, University of Tasmania
Dev Ranmuthugala
Australian Department of Defence
Marcus Bowles
Deakin University

Abstract

The objective of this research was to investigate the factors of assessment (task, criteria, etc.) that students undergoing authentic assessment perceived significant towards their academic achievement. This project advanced past research by the authors of this paper that found seafarer students' academic achievement was significantly higher in a formatively implemented authentic assessment compared with a summative traditional assessment. The academic achievement (assessment scores) was based on the students' performance in analysing information presented in a real-world context (authentic assessment) as opposed to the analysed of information presented devoid of context (traditional assessment). Using the students undergoing the authentic assessment, this project correlated their perceptions of authenticity for factors of assessment to their scores in the associated task. Stage 1 focused on deriving the factors conceptually, based on which a perception survey questionnaire was designed. Following the collection of student responses through the survey, a correlational analysis was conducted between student perceptions and their scores. Stage 2 extracted new factors through a factor analysis. Using the survey data (but for the new factors), an additional correlational analysis was conducted to confirm findings. Both stages of investigation found that the factor of transparency of criteria was a significant predictor of the students' academic achievement.

Keywords: authenticity, assessment, student, perceptions, seafarer, factor analysis

Introduction

Endedijk and Vermunt (2013) and Beyaztas and Senemoglu (2015) noted that a focus on measuring decontextualised memorization and understanding of content, and not the integration or application of knowledge (Biggs and Tang 2011; Kearney 2012) is a common issue leading to superficial approaches to learning in many higher education systems worldwide. Similar issues were noted by researchers (Maringa 2015; Emad and Roth 2007; AMC 2011) in the area of seafarer (persons employed on ships) education and training. Empirical evidence in past research suggested that seafarer students tended to disengage with learning and assessment when traditional assessment tasks required them to construct responses purely based on the analysis of information presented devoid of context making them rely solely on their ability to regurgitate memorised information, for example through oral examinations, written assignments, and multiple-choice questions (MCQs). Relying on memorisation is a lower-level cognitive ability, lapses in which lead to unintentional errors and lower academic achievement in assessments (Wiggins 1989).

Past research (Brawley 2009; Schneider, Krajcik, Marx, and Soloway 2001; Thomas 2000; Leon and Elias 1998; Gallagher, Stepien, and Rosenthal 1992) empirically proved that academic achievement was higher for students undergoing authentic assessments as compared with traditional assessments. In authentic assessments, students construct responses based on the assimilation, integration, and analysis of critical information presented in a real-world context (Wiggins 1989). Hence, authentic assessments appear as a model that integrates knowledge and skills acquired in the classrooms with employment, replicating the tasks and performance standards typically faced by professionals in the world of work (Villarroel, Bloxham, Bruna, et al. 2018), making it suitable for implementation in the vocational-based seafarer education and training. However, an extensive literature review by the authors of this paper revealed that there was a global absence of evidence regarding the impact of authentic assessment in seafarer education (Ghosh, Bowles, Ranmuthugala, and Brooks 2016; 2017). To address this gap, the authors investigated [Ghosh et al. To be published] seafarer students' academic achievement (measured through their assessment scores) in authentic assessment as compared with traditional assessments. However, past researchers (Bailey 1998, 205; Law and Eckes 1995; Dikli 2003, 16; Abeywickrama 2012) described traditional assessments as a 'one-shot' or single-occasion tests that are implemented at the end of the learning period in a summative manner. Since, authentic assessments were characterised with providing students with more than one opportunity (Wiggins 1989; Gulikers 2006), the authors also distinguished the two assessments on the basis of the implementation as well. The traditional assessment was implemented in a summative format while the authentic assessment was implemented formatively.

Two separate and independent student groups were used as the 'control' (traditional) and 'treatment' (authentic) group in a selected unit of learning delivered within the Bachelor of Nautical Science degree programme conducted at the Australian Maritime College (AMC) of the University of Tasmania (UTAS). The traditional assessment comprised of two case study scenarios to which the students constructed their responses to essay-type questions, based on their analysis of the scenarios presented devoid of the context. In comparison, a new cohort of students enrolled in the same unit in the following semester was assessed authentically using the same case studies. Although, the authentically assessed students also

provided written responses to the same essay-type questions, the authentic assessment differed from the traditional assessment by providing a real-world authentic context to the assessment task through simulations and practical demonstrations of the same case study scenarios, enacted by AMC staff. For example, one case study that described ship staff abandoning the ship using a liferaft during an on-board fire was demonstrated by emergency response staff at AMC's training pool. The pool was equipped with facilities to launch a real liferaft in simulated waves, strong winds, darkness, rain, and smoke. The simulation also included sounding of the emergency alarms with staff playing the role of panicking seafarers jumping into the pool to replicate a possible emergency. In comparison to the authentic assessment, students assessed traditionally relied only on their imagination and experience to visualise the described scenarios.

The authentic assessments were formative in nature and held on two different days, three weeks apart. The second authentic task was implemented once the students received feedback on their performance from the first authentic task. In comparison to the authentic assessment, the traditional assessment was summative in nature and both case studies were implemented at the assessment. However, the duration of the authentic assessment (combined) was the same as that of the traditional assessment. The assessment details and rubric were provided to both the student groups at the beginning of the semester. To avoid the introduction of additional variables, the unit, learning content, lecture delivery methods, lecturer, assessment rubric, total duration of the assessment, assessment questions, and response methods were kept constant.

Findings of the past research conducted by the authors (Ghosh et al. To be published) confirmed that seafarer students' academic achievement was significantly higher in the formative authentic assessment when compared with the summative traditional assessment. Although, in past research work, higher academic achievement was attributed to the 'authentic' design of the assessment and the formative nature of its implementation, further research was required to investigate the factors of assessment that the students may have perceived significant and influenced their perception of authenticity in assessment leading to higher academic achievement. These factors will provide guidance to assessors in the designed authentic assessment with the aim of improving scores and the resulting academic achievement. Hence, using the same but independent sample of authentically assessed students, the research presented in this paper investigated student perceptions of authenticity in assessment to reveal the factors of assessment that correlated significantly with their academic achievement.

As a result, the following research question (RQ) was developed:

RQ: What is the correlation between seafarer students' perception of authenticity in assessment and their academic achievement in the associated assessment task?

The developed RQ enabled the development of the following research variables:

- independent variable: Perceptions of authenticity in assessment; and
- dependent variable: Students' academic achievement.

This research identified seafarer students' 'perception of authenticity in assessment' as the independent variable. The term 'authenticity' in this regard referred to the characteristics (e.g. setting assessment tasks in real-world contexts) of the authentic assessment that students may perceive significant towards the outcomes of: higher student engagement; ability to transfer skills to different contexts; contextual and multiple evidence of competence; and valid (relevant to workplace) and reliable (multiple and consistent) student performance (Ghosh et al. 2017). The defining characteristics of authentic assessment that lead to the aforementioned outcomes are explained in Ghosh et al. (2017); and summarised in Table 1. Subsequently, the key words (bold in Table 1) in the defining characteristics of authentic assessment were used to conceptually develop the factors of assessment (task, context, criteria, etc.). The development of the factors is also shown in Table 1.

Table 1: Defining independent variable to provide conceptually developed factors of assessment for measuring seafarer students' perception of authenticity.

Independent variable	**Defining characteristics**	**Conceptually developed factors of assessment derived from keywords in the defining characteristics**
Perception of 'authenticity' in assessment	Assessment outcomes: Higher student engagement; Ability to transfer skills to different contexts; Contextual and multiple evidence of competence; Valid and reliable student performance	
Authentic assessment outcomes:		
Higher student engagement	Setting assessment **tasks** in real-world **contexts**; Assessment tasks should be **relevant to the workplace;** Assessment's emphasis on active **construction of knowledge**; Performance **criteria** should reflect workplace needs and be provided beforehand to show **transparency**; **Multiple opportunities** for students to improve learning based on **feedback** on learning achieved	Task; Context; Relevance to the workplace; Construction of knowledge; Criteria; Transparency of criteria; Multiple opportunity based on feedback
Ability to transfer skills to different contexts	Setting assessment **tasks** in real-world **context**; Students using **feedback** to identify and fill gaps in competence through **multiple opportunities**	Task; Context; Multiple opportunity based on feedback
Contextual and multiple evidence of competence	Students provided with **multiple opportunities** to improve learning based on previous **feedback**	Multiple opportunity based on feedback
Valid and reliable student performance	Assessment tasks should be **relevant to workplace**; **Multiple opportunities** to improve learning based on previous feedback	Relevance to the workplace; Multiple opportunity based on feedback

Based on the conceptually developed factors (Table 1), this project constructed a questionnaire which was used to obtain student responses regarding their perception of authenticity in assessment. In Stage 1, the perceptions of authenticity for the conceptually developed factors were correlated to the dependent variable of students' academic achievement (defined by their composite numeric scores obtained in the authentic assessment tasks). Stage 2 extracted new factors of assessment through a factor analysis. Using the student responses from the perception survey, an additional correlational analysis was conducted between students' perception of authenticity for the new factors of assessment and their scores in the authentic assessment.

Both stages of investigation revealed significant findings towards the design of authentic assessments for higher academic achievement of students.

Research methodology

Questionnaire design

This paper used a questionnaire to measure seafarer students' perception of authenticity in assessment. To develop the questionnaire, past research in the area of authentic assessment was scanned to investigate if existing published questionnaires and/or items could be used for the purpose. Additionally, an internet search was conducted for the same purpose. The final survey document developed for this research used all the questions from Gulikers (2006) to form Questions 5—27. Since Guliker's (2006) questionnaire was developed for social work students, the word 'social worker' was replaced with the word 'seafarer' in the questionnaire developed for this project. One question was adopted from the National Survey of Student Engagement (NSSE) to form Question 28a—28e. Two questions were devised by the authors of this paper to form Questions 29 and 30a—30b. The first four questions enquired student demographic details. Questions 5—27 and 29—30 were scored on a 5-point Likert scale ranging from 1 (totally disagree) to 5 (totally agree). Only Question 28 was scored on a 4-point Likert scale ranging from 1 (very little) to 4 (very much). The Likert scale was reverse coded for negatively worded questions (i.e. Questions 10, 11, 18, 23, 26, 28a). Question 30a required a response on the nominal scale of 'Yes' or 'No'.

Validity and reliability of the questionnaire

Since the questionnaire constructed for this research was mainly drawn (barring three questions) from Gulikers (2006), it initially derived its validity and reliability from the values published by that author. According to Gulikers (2006), all scales of the survey had a reasonable internal consistency, shown in Cronbach's alpha ranging from 0.63 to 0.83. The Cronbach's alpha for the survey used in this research had a value ranging from 0.69 to 0.75. The questionnaire was also reviewed through a pilot survey by 12 fellow academics and researchers within the AMC, where the research was conducted. The pilot survey respondents suggested defining the terms 'context', 'criteria', 'oriented', 'under-graduate', post-graduate', and 'output' for the students. The respondents also suggested excluding the demographic

question enquiring the age of the students and including the question related to the educational qualifications.

Data collection

The survey was administered on completion of the authentic assessments for the treatment group. A general announcement was made in class and an email was sent inviting students to participate in the survey. A minimal risk ethics application approval, constituting ethical clearance by the Tasmania Social Sciences Human Research Ethics Committee, was obtained for this research project. Participants were reassured that the data would be anonymised and that their contribution would be confidential. Students were free to withdraw at any time from the study.

Sampling considerations and response rate

The sampling technique used in this research was based on convenience sampling that relies on opportunity and participant accessibility and is used when the study population is large, and the research is unable to test every individual (Clark 2014; Robson 2011). Thus, participants for this research were two separate groups of seafarer students drawn from the Bachelor of Nautical Science programme at AMC enrolled in the selected unit.

A key consideration while sampling was to ensure that the treatment group was comprised of randomly assigned students in which each participant had an equal chance of being chosen based only on the sequence of enrolment in the individual semesters. The groups were not sorted based on any other pre-determined characteristics, such as qualifications, academic ability, age or work experience that may have impacted the outcomes of this research. This ensured that the relationship between the two variables remained the same in all segments of the sample, which is essential for correlational research (Graziano and Raulin 2000). Moreover, in correlational research the coefficient of determination (r^2) that allows us to estimate how useful the relationship between the dependent and independent variables might be in a prediction (and is a measure of effect size), should be considered significant only if the minimum sample size is 30 (Suresh and Chandrashekara 2012; Lodico, Spaulding, and Voegtle 2010; Blondy 2007; Graziano and Raulin 2000). This research, thus, exceeded the recommended minimum sample size.

Although 102 students were asked to respond to the survey, only 98 students participated in the study. Out of the 98 respondents, only 93 surveys were usable for analysis, as 5 surveys were discarded due to incomplete/absent responses.

Data analysis

The correlation analysis was conducted in two stages using the statistical software package SPSS 23.

Stage 1: Correlation analysis between students' perception of authenticity in assessment (for factors derived conceptually) and their scores

The questionnaire statements were categorised under the conceptually developed factors of assessment (task, context, criteria, etc.) as determined in Table 1. Questions categorised under a common factor were subjected to an inter-reliability analysis (Cronbach's alpha) to ensure that they were significantly correlated to each other. This is detailed in Table 2.

Table 2: Survey questions categorised under conceptually developed factors of assessment; and their inter-reliability values.

Question number	Question statement	Factors of assessment	Cronbach's alpha
5	This assessment was oriented to my future profession of a seafarer.	Relevance to the workplace	0.840
6	This assessment was clearly directed to my professional requirements.		
7	This assessment prepared me for my future profession.		
15	This way of assessing is an effective way of assessing professional skills.		
16	This way of assessing fits well with the seafarer's profession.		
17	The output that I had to produce in this assessment is part of the seafarer's job.		
18	The output that was evaluated in this assessment is different from what is being evaluated in practice.		
19	The result that I had to produce in this assessment is something that a real seafarer also has to produce in practice.		
27	In this assessment, both knowledge and professional skills were important.		
8	The task of the assessment resembled the task of a real seafarer.	Task	0.478
9	The task of this assessment was an important part of the seafarer profession.		
10	The task of this assessment differed from the tasks of a real seafarer.		
11	The context in which I had to perform the assessment was fake.	Context	0.650
12	The context in which I had to perform the assessment looked like a seafarer's workplace.		
13	The context in which I had to perform the assessment looked just like the real world.		
14	The context in which I had to perform the assessment was realistic.		
20	The criteria resembled the criteria that I have to meet in practice.	Criteria	0.547
21	The criteria that I had to meet in this assessment resembled the criteria used in practice.		
22	In this assessment, I was evaluated on criteria important for the seafarer's profession.		
23	In this assessment, I was evaluated on things that I never have to use in real profession practice.		
24	The criteria that I had to meet in this assessment were clear enough.	Transparency of criteria	0.763
25	Before I started the assessment, it was clear to me what was expected of me.		
26	It was hard to find out what was expected of me in this assessment.		
28	The following requirements of the assessment helped me to improve my score:	Construction of knowledge	0.540
28a	Memorising course material		
28b	Applying facts, theories, or methods to practical problems or new situations		
28c	Analysing an idea, experience, or line of reasoning in depth by examining its parts		
28d	Evaluating a point of view, decision, or information source		
28e	Forming a new idea or understanding from various pieces of information		
29	The feedback provided in this assessment helped me to identify the strengths and weaknesses in my learning.	Multiple opportunity based on feedback	0.697
30a	This assessment provided more than one (1) opportunity to improve my score.		
30b	If 'Yes', the feedback provided on my first performance helped me to improve my assessment score in the next performance.		

For the purposes of this paper, a Cronbach's alpha value of greater than 0.70 (Tavakol and Dennick 2011) was considered statistically significant for reporting. Table 2 showed that an inter-reliability analysis of the categorised survey questions revealed an acceptable Cronbach's alpha value (0.70 or greater) for only two factors of assessment, i.e. relevance to workplace and transparency of criteria. Since, an acceptable value of Cronbach's alpha was found for a low number, i.e. only two factors, a correlation analysis between seafarer students' perception of authenticity in authentic assessment for all the developed factors and their scores in the associated assessment task was conducted. The correlation between the variables (perception of authenticity and scores) was considered significant if the correlation coefficient (R) value was higher than 0.25 (Clark 2014). The findings of the correlation analysis conducted in stage 1 are discussed in the 'Results' section later on in this paper.

Stage 2: Correlation analysis between students' perception of authenticity in assessment (for factors extracted through factor analysis) and their scores

Since, the majority of the conceptually developed factors of assessment (except transparency of criteria and relevance to workplace) had a low value (less than 0.70) of Cronbach's alpha, a factor analysis to statistically develop new factors of assessment was conducted. Next, a factor analysis to remove multicollinearity and extract factors that are relatively independent of one another was conducted. Factors extracted after the exploratory factor analysis (EFA) and their contribution towards explaining the variance in data are shown in Appendix 1.

Appendix 2 reports the loading of the survey questions under the factors derived from the factor analysis. The questions loaded cleanly (without overlap) under the seven factors. The construction of knowledge questions (28b—28d) clustered in Factor 2, the 'context' questions (question 12—14) item in Factor 4, the 'transparency of criteria' questions (24—26) in Factor 5, and the 'multiple opportunity' questions (29—30b) in Factor 7. Hence, these factors retained the original titles. The questions that were reverse coded clustered in Factor 6, which was therefore titled irrelevant to the profession.

Conversely, the questions related to the conceptually developed factors of relevance to the workplace, task and criteria did not cluster in the expected way; and loaded unevenly (split loading) in Factors 1 and 3. Although a limitation of factor analysis is that factor names may not accurately reflect the variables within the factor, especially in the case of split loadings (Yong and Pearce 2013), this research used the factor naming technique suggested by Neill (2008). Neill advocated for using the majority of the loading items for naming each factor. The items in Factors 1 and 3 were reviewed to provide meaningful names for the extracted factors based on the top loadings for each factor. Additionally, each factor was subjected to an inter-reliability analysis (Cronbach's alpha) to verify if the values were greater than 0.70. Table 3 details the survey question numbers with their factor loadings, together with the factors titles, and the Cronbach's value of inter-reliability analysis.

Table 3: Factors extracted using factor analysis: Categorised survey questions, titles, and inter-reliability values.

Factor	Survey questions	Factor title	Cronbach's alpha
Factor 1	5, 6, 7, 8, 9, 17, 22	Relevance to the profession	0.865
Factor 2	28b, 28c, 28d, 28e	Construction of knowledge	0.806
Factor 3	15, 16, 19, 20, 21, 27	Assessing competence to job-relevant criteria	0.868
Factor 4	12, 13, 14	Context	0.732
Factor 5	24, 25, 26	Transparency of criteria	0.763
Factor 6	10, 11, 23	Irrelevant to the profession	0.616
Factor 7	29, 30b	Multiple opportunity	0.697

Based on the inter-reliability values of Cronbach's alpha, Table 7 revealed that the factor analysis extracted five factors with an acceptable value of more than 0.70. Factors 6 and 7 were rejected due to low Cronbach's alpha values of less than 0.70. The selected factors (1—5) cumulatively explained 60% of the variance in the data, which was considered significant (Williams, Brown and Onsman 2010) for further correlation and regression analysis. Thus, stage 2 investigated the correlation between seafarer students' perceptions of authenticity for the new factors (1—5) of assessment extracted through factor analysis and their scores in the associated assessment task. The correlation between the variables (perception of authenticity and scores) was considered significant if the correlation coefficient (R) value was higher than 0.25 (Clark 2014). The findings of the correlation analysis conducted in stage 2 are discussed in the 'Results' section below.

Results

The results of the RQ is summarised for each stage of investigation in Table 4.

Table 4: Results summary.

RQ	Stage	Results summary
RQ1	Stage 1	Significant correlation found between seafarer students' perception of authenticity for the conceptually developed factor of transparency in criteria and their scores in the authentic assessment. Transparency of criteria was also found to be a significant predictor of student scores in the authentic assessment.
RQ1	Stage 2	Significant correlation found between seafarer students' perception of authenticity for Factors 2 (construction of knowledge) and 5 (transparency in criteria) extracted through a factor analysis and their scores in the authentic assessment. Factor 5 was also found to be significant predictor of student scores in the authentic assessment.

Reporting of the results below is organized by each stage of data analysis.

Stage 1

The R-values for the correlation between the students' perception of authenticity (for the conceptually developed factors) in authentic assessment and their scores in the associated assessment task are detailed in Table 5.

Table 5: R-values of student perceptions of authenticity (conceptually developed factors) in authentic assessment and their scores in the associated assessment task.

	Relevance	*Task*	*Criteria*	*Trans Cri*	*Context*	*Constn Kn*	*Mul Opp*	*Totl Score*
Relevance	1							
Task	0.7151807	1						
Criteria	0.7434481	0.557436508	1					
Trans Cri	0.5277081	0.357781976	0.57022824	1				
Context	0.5614482	0.595435346	0.430059512	0.252956709	1			
Constn Kn	0.6740382	0.481201314	0.45849741	0.497534833	0.417284643	1		
Mul Opp	0.4850502	0.302732244	0.367019161	0.393821471	0.256092255	0.448560931	1	
Totl Score	0.171476	0.194456237	0.136400319	**0.31225956**	0.109658853	0.188522718	-0.022574955	1

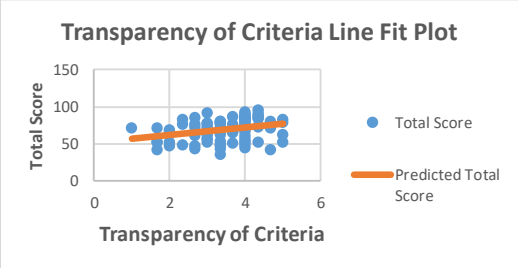
The R-values in Table 5 showed significant correlation (R-value higher than 0.25 was outlined in bold) between students' perception of authenticity for the factor transparency of criteria and their scores in the authentic assessment. Using the significantly correlated factor (transparency of criteria) and the scores in the authentic assessment, a linear regression

analysis based on the recommended (Sarkar, Keskin, and Unver 2011) confidence level of 95% (or p-value 0.05 or less) was conducted. Although confidence levels can be represented as 90%, 95%, 99% or any percentage (between 0 and 100%), the authors of this paper chose the most commonly used confidence level of 95% (Tan and Tan 2010). The findings of the regression analysis are detailed in Figure 1.

Figure 1: Regression analysis of seafarer students' perception of authenticity in transparency of criteria and their scores in authentic assessment.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.3122596
R Square	0.097506
Adjusted R Square	0.0875885
Standard Error	13.967688
Observations	93



ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1918.128195	1918.128195	9.831698956	0.002310037
Residual	91	17753.76428	195.0963108		
Total	92	19671.89247			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	51.299776	6.090641979	8.422720581	5.07168E-13	39.20146347	63.39808764	39.20146347	63.39808764
Trans Cri	5.2648653	1.679086156	3.135554011	**0.002310037**	1.929566951	8.600163712	1.929566951	8.600163712

The bold p-value (less than 0.05) of the factor transparency of criteria, as shown in Figure 1, revealed the factor to be a significant predictor of student scores in authentic assessment. However, this finding was based on a relatively lower value (8.8%) of the adjusted R-square.

Stage 2

The R-values for the correlation between the students' perception of authenticity for the factors of assessment (extracted through factor analysis) and their scores in authentic assessment are detailed in Table 6.

Table 6: R-values of student perceptions of authenticity (factors extracted through factor analysis) in authentic assessment and their scores in the associated assessment task.

	Factor 1	*Factor 2*	*Factor 3*	*Factor 4*	*Factor 5*	*Totl Score*
Factor 1	1					
Factor 2	0.51256627	1				
Factor 3	0.778340265	0.608966905	1			
Factor 4	0.464486019	0.349940441	0.470277922	1		
Factor 5	0.418538139	0.503530125	0.566614622	0.190884736	1	
Totl Score	0.160202988	**0.296878226**	0.169033928	0.030550485	**0.31225956**	1

The R-values in Table 6 showed significant correlation (R-value higher than 0.25 was outlined in bold) between students' perception of authenticity for Factors 2 and 5 and their scores in the authentic assessment. Using the significantly correlated factors (2 and 5) and the scores in the authentic assessment, a multiple regression analysis based on the recommended (Sarkar et al. 2011) confidence level of 95% (or p-value 0.05 or less) was conducted. The findings of the regression analysis are detailed in Figure 2.

Figure 2: Regression analysis of seafarer students' perception regarding authenticity in Factors 2 and 5 and their scores in authentic assessment.

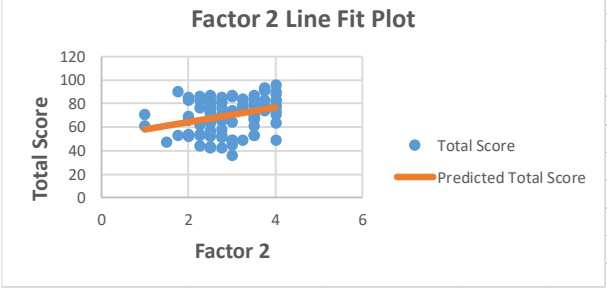
SUMMARY OUTPUT

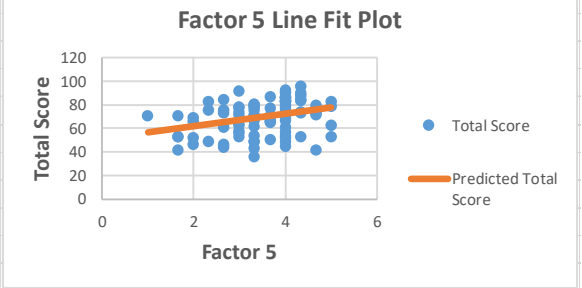
Regression Statistics	
Multiple R	0.351611748
R Square	0.123630821
Adjusted R Square	0.104155951
Standard Error	13.84029554
Observations	93

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	2432.052221	1216.026	6.348223	0.002635842
Residual	90	17239.84025	191.5538		
Total	92	19671.89247			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	45.50989075	6.994083563	6.506913	4.23E-09	31.61492299	59.4048585	31.61492299	59.4048585
Factor 2	3.868048422	2.361499213	1.637963	0.104921	-0.823481943	8.559578786	-0.823481943	8.559578786
Factor 5	3.676608053	1.925711831	1.90922	**0.05942**	-0.149154615	7.502370721	-0.149154615	7.502370721





The bold p-value (less than 0.05) of the factor transparency of criteria, as shown in Figure 2, revealed the factor to be a significant predictor of student scores in authentic assessment. However, this finding was based on a relatively lower value (10.4%) of the adjusted R-square.

Discussion

Transparency of assessment criteria in authentic assessment

Transparency of assessment criteria is essential for learning (Biggs and Tang 2011; Reddy 2007; Wiggins 1989) and providing the criteria at the beginning of the learning period (thus making the assessment transparent) is an essential and key requirement for authentic assessment (Villaroel et al. 2018). Findings of this research project (as shown in Figures 1 and 2) confirmed that the factor transparency of criteria is essential for learning since students had significantly higher achievement when they found the assessment criteria to be transparent. Having the assessment criteria (detailing standards of performance) beforehand provided a roadmap of the subject to be learned, while allowing students to construct the understanding of the topic. Students were also able to reflect on their learning and carry out self-assessments of their thinking and practices towards achievement of the required standards. This finding corroborated the findings of the past research (Gulikers 2006; Jonsson 2008) when transparency of assessment criteria enhanced student achievement in authentic assessment. Similar to the research presented in this paper, Gulikers (2006) found the transparency of assessment criteria to be the strongest influence on social work students' learning and their skills development out of other factors such as task and context. Jonsson (2008) focused only on the correlation between transparency of assessment criteria and student scores and revealed that increasing transparency of criteria improved teacher students' performances.

The significance of transparency of criteria on student achievement was also concurred by Hattie (2009; 2007). To ascertain the major influence on student achievement, Hattie (2009; 2007) synthesised more than 800 meta-analyses in education and found that making the criteria more explicit leads to skills improvement since students become more aware of what constitutes a successful performance. Clarity in expectations engages students in the task, which further increases the chance of enhancing their achievement. Hattie (2009; 2007) also argued that without transparency in assessment criteria, providing students with the feedback on their performance is devoid of context. Feedback directed to the transparent criteria enables students to reduce the gap between their current level of competence and the expected level. Well-directed feedback can then be used by students to adjust their learning strategies towards higher achievement (Hattie 2009; 2007).

Providing students with feedback in authentic assessment

The positive impact of providing students with feedback on their academic achievement was reaffirmed empirically in this paper, thus, advancing past research (to be published) by the authors. The findings of the past research by the authors evidenced that higher academic

achievement in authentic assessment was not only due to the authentic element of the assessment but also attributed to the formative nature of its' implementation. The correlation study in this paper confirmed that the group of seafarer students that had significantly higher academic achievement in the authentic assessment perceived the factor transparency of criteria to be the most significant predictor of their achievement (Figures 1 and 2). This also indicated that the seafarer students, who underwent formative authentic assessment were able to improve their performance in the second assessment task resulting in an improved academic achievement using the feedback provided to them on their performance in the first task. The feedback provided enabled the students to conduct a self-assessment of their existing knowledge and skills using the assessment criteria provided at the beginning of the learning period. The students, then, adopted learning strategies towards obtaining higher academic achievement in the second authentic assessment task. Higher academic achievement also confirmed that feedback provided on student performance in authentic assessments allows them to improve on it which is not the case with traditional assessments. The latter was evidenced in an earlier study by Huang (2017) in which law students commented on the lack of adequate feedback on task performance and ways to improve on it when traditional assessments like multiple-choice questions were used to assess them.

Contribution in seafarer education and training

Through the correlation analysis, the research presented in this paper revealed the factors of assessment that had a significant relationship with seafarer students' academic achievement (Figures 1 and 2). The factors that correlated significantly not only explained higher student achievement in authentic assessment, but also provided valuable insights into seafarer assessment design, where the inclusion of the highly correlated factors in assessments may lead to improved student scores in the future. Hence, this research makes a valuable contribution through the collection of empirical evidence towards the impact of authentic pedagogical practices in seafarer education. This is based on an extensive literature review and past research (Ghosh et al. 2016; 2017), which revealed a global absence of research on authentic assessment in the area of seafarer education.

Enhancing generalisability of research findings

Past research findings that provided empirical evidence on the impact of authentic pedagogical practices on outcomes outlined in this paper (student engagement, ability to transfer skills, etc.) using qualitative methodology, were based on a low sample. For example, past research investigating the impact of authentic pedagogy on student engagement was based on 6 students (Richards Perry 2011), 11 students (Quartuch 2011), 4 students (Findlay 2013), and 10 students (Morrissey, 2014). Although Morrissey (2014) justified the use of qualitative methodology (interviews) on a low sample as a means to gain an in-depth understanding of the concerned research phenomena, the findings did not support representativeness or generalisability. Hence, this research contributed towards enhancing the generalisability of research findings in the area of authentic assessment by using a quantitative questionnaire and obtained data from a relatively larger sample of 93 students.

Significance of construction of knowledge in authentic assessment

In stage 2, Factor 2 (construction of knowledge) also significantly correlated to the student scores in authentic assessment. However, a further regression analysis, assuming a 95% confidence level, did not find the factor to be a significant predictor of scores. If this paper had assumed a 90% confidence level, Factor 2 also would have been considered a significant predictor of students' academic achievement. The choice of whether to use a 90 or 95% confidence interval is somewhat arbitrary (Tan and Tan 2010), and the 95% confidence level for this research was chosen due to its common use. However, this should not diminish the value of the factor of construction of knowledge and hence, should be included in designing authentic assessment for students.

Low value of adjusted R-square

The findings of the regression analysis presented in this paper, are based on a relatively low value (8.8% in stage 1 and 10.4% in stage 2) of adjusted R-square. The adjusted R-square value focuses on explaining the observed variation in the dependent variable due to the independent variable (Lukacs, Burnham, and Anderson 2010). This implied, that the significant factor (transparency of criteria) in this paper, although important, did not explain the majority of the variance in the student scores. This was also evidenced by the fact that Factor 1 accounted for the majority of the variance (38.5%) and, did not correlate significantly to the scores. Hence, it was a possibility that the correlation and regression model adopted in this paper may not have included important factors of assessment before measuring the independent variable of perception of authenticity in assessment. For example, factors of assessment such as collaborative assessment (Ashford-Rowe, Herrington, and Brown 2014; Gulikers 2006), student ownership of task design (Gulikers 2006), completion of task and collating of evidence competence by students over a sustained period (Morrissey 2014); and presentation of student work to an audience (Herrington 1997) were rejected at a theoretical level due to the following reasons:

- Collaborative assessment was rejected since the research by Gulikers (2006) revealed that students and teachers rated this factor (described as 'social context') as the least important dimension of authentic assessment. Moreover, demonstrating individual competence in the units of learning is essential for seafarer certification (International Maritime Organization (IMO) 2011).
- Factors such as collaborative assessment, student ownership of task design, and completion of task over a sustained period of time were also rejected to avoid plagiarism in student work. This research required seafarer students to complete the assessment task under the supervision of externally employed invigilators. The factors were also rejected since inclusion of these factors in the assessment design would have created uncontrolled additional variables (e.g. variation in student groups, variation in task design and variation in time taken to complete task) other than the authentic design that would have impacted student performance.

- The factor requiring presentation of student work to an audience was rejected since it was incongruous to the nature of the assessment task developed for this paper.

The relatively low value of adjusted R-square may have also resulted from the use of the quantitative survey to measure student perceptions. This is because the use of Likert scales in the quantitative survey may have limited the students from outlining, describing, and adequately conveying the other factors of authentic assessment that they perceive to be significant towards obtaining a higher academic achievement. Instead, the students were compelled to choose the significant factors amongst the choices provided through the survey which may have led to an inadvertent omission of factors. This was also evidenced by the perception study by Gulikers (2006) in which, the quantitative data did not reveal an overall differing perception of authenticity in task, but the qualitative investigation revealed otherwise.

Goodwin and Leech (2006) recommended examining the variability in the data (dependent and independent variable) if the resulting correlation was lower than expected. Lack of variability (indicated through low values of standard deviation) lowers the correlation value between variables (Goodwin and Leech 2006). To examine the variability, this research calculated the standard deviation values for the student survey responses for perception in authenticity (independent variable) and the composite student scores (dependent variable). The standard deviation for student scores was 14.6 (mean score 69.8/100; minimum 36/100; maximum 96/100). The standard deviation, thus, indicated a relatively low value, which may have contributed to the lower correlation between the variables. Similar to the dependent variable, the standard deviation values of the student responses to the perception survey (as shown in Appendix 3) had relatively low values, which may also have contributed to the lower correlation.

Lack of variation in student scores indicated evenness in student performance. This may imply that the evenness in performance may have resulted due to the transparency in assessment criteria that provided all students with the same guidelines to obtain higher academic achievement. This argument is based on past researchers (Black and William 1998a, 1998b; Sadler and Good 2006; Jonsson 2008), who claimed that transparency in criteria is not only an effective means to improve performance but also a provider of equality in academic achievement. The researchers argued that in studies characterised by formative assessments and transparent criteria, the difference in student achievement between high- and low-performing students is typically reduced.

Conclusion

Past research [To be published] by the authors found that seafarer students' academic achievement was significantly higher in formatively implemented authentic assessment in which students constructed responses based on the assimilation, integration, and analysis of information presented in real-world settings. This was opposed to a summative traditional assessment that focused on students constructing responses based on memorisation and regurgitation of information. Building on past research (Ghosh et al. To be published), the authors investigated factors of authentic assessment (task, context, transparency of criteria,

etc.) that correlated significantly to higher academic achievement (measured using scores obtained in the assessment tasks). Findings derived through factor analysis confirmed that the factor transparency of criteria correlated significantly with student scores. This finding confirmed that providing students with assessment criteria at the beginning of the learning period provides them with clear indications on standards of performance expected in the assessment tasks. Using the feedback provided on their performances in formatively implemented authentic assessment tasks, the students conduct a self-assessment of their learning. Once the gaps in their knowledge and skills are recognized, the students focus on aspects of learning that will improve their performance and overall scores, making them autonomous learners and eventually, skilled professionals.

Although empirical evidence showing higher student academic achievement in authentic assessment (as compared with traditional assessment) has been presented in the past, this research contributed to the existing evidence through the data collected in the area of seafarer education and training. Such evidence was globally missing. The empirical data collected from a relatively large sample in this project also enhanced the generalisability of findings evidenced towards the impact of authentic pedagogical practices on assessment outcomes such as student engagement, ability to transfer skills, etc.

One may argue that a key limitation of this paper was that the findings of this project were based on a relatively low value of adjusted R- square. The adjusted R square value focuses on explaining the observed variation in the dependent variable due to the independent variable. However, the focus of this paper was not to explain variation, but to find an association through correlation between the independent variable (perception of authenticity) and dependent variable (scores). In this context, the adjusted R-square value was irrelevant; and the low R- square value with statistically significant parameters was more valuable than a high R-square value accompanied with statistically insignificant parameters. This research acknowledged that a limitation of this project resulted due to the use of the quantitative methodology adopted to enhance generalisability of findings. The quantitative methodology used a survey based on a Likert scale that limited the response of the seafarer students to a perception survey. Hence, certain variables (collaborative assessment and student ownership of task design) may have been rejected at a theoretical level and intentionally omitted from the data analysis model used in this project. Therefore, future research will investigate seafarer students' perceptions with regard to authenticity in assessments when compared to the actual workplace through the use of qualitative methodologies, such as interviews, focus groups etc., Although certain variables were excluded, this research uncovered significant factors of assessment which, if included, in the design of the assessment will guide authentically assessed students towards higher academic achievement.

APPENDICES

APPENDIX 1: EXTENT OF VALIDITY AND RELIABILITY TESTING OF AUTHENTIC ASSESSMENT WHEN IMPLEMENTED USING RUBRICS (FIRST LITERATURE REVIEW)

Author	Context of Implementation	Research Aims	Validity	Reliability
(Einbender and Wood, 1995)	School Students	To improve schooling and all aspects of educational enterprise through authentic assessment and practice.	None	None
(Chance, 1997)	University Students (Statistics Course)	To determine the goals and experiences of the University with authentic assessment techniques and develop strategies for effective implementation.	None	None
(Todorov and Brousseau, 1998)	School Students	To determine evidence of achievement of content standards through authentic assessment	None	None
(Emery, 2001)	School Students	To determine if student performance can be improved using authentic assessment	None	None
(Vendlinski et al., 2002)	School Students	To determine a correlation between different measures of student understanding through authentic assessment	None	None
(Bell and Bell, 2003)	Professional (Teachers)	To assist teachers to create and/or improve methods of assessment for students.	None	None
(Moon et al., 2005)	School Students	To determine quantifiable information about student learning and instruction process.	Content	Inter-rater
(Oh et al., 2005)	University Students (Sciences)	To develop and use valid and reliable authentic assessment measures.	Content	Inter-rater; Internal consistency
(Mallet, 2006)	University Students (Mathematics and Engineering)	To provide the students with authentic activities preparing them for their future work	None	None
(Johnson, 2007)	School Students	To compare student achievement scores on authentic assessment with that on traditional assessments	Face; Content	Internal - Consistency
(Olfos and Zulantay, 2007)	School Students	To improve the validity and reliability of the web-based authentic assessment system	Concurrent	Internal-Consistency
(Jonsson, 2008)	University Students (Teacher Education)	To assess student performance and self-assessment skills of students in authentic assessment	Face; Construct	Internal - Consistency; Inter-rater; Rank Correlation

Appendix 1

(Diller and Phelps, 2008)	University Students (Information Literacy)	To assess the effectiveness of the course program through authentic assessment	Validity demonstrated through reliability tests	Internal – Consistency
(Brawley, 2009)	School Students	To determine if authentic assessment requires higher-order thinking skills than traditional assessments	None	None
(Cassidy, 2009)	University Students (Teacher Education)	To measure teacher effectiveness and student achievement through authentic assessment scores	Convergent (Same as Construct)	None
(Koh and Luke, 2009)	School Students	To determine if teacher professional development in authentic intellectual assessment task design can contribute to the improvement of student learning and performance	None	Inter-rater
(Taylor, 2011)	School Students	To measure achievement of learning objectives through interdisciplinary authentic assessment	None	Inter-rater
(Azim and Khan, 2012)	School Students	To assess students' knowledge, higher-order skills, and performance through authentic assessment	None	None
(Kearney, 2012)	University Students (Teacher Education)	To develop an authentic self- and peer-assessment for learning	None	None
(Lang II, 2012)	University Students (Teacher Education)	To compare validity between authentic assessment and traditional professional examination	Content	None
(Blackburn and Kelsey, 2013)	School Students	To assess student performance in authentic assessment	None	None
(Fatonah et al., 2013)	School Students	To assess student performance in a proposed authentic assessment model	Content	1) Inter-rater 2)Instrument
(McCarthy, 2013)	University Students (Business)	To use as a self-assessment and feedback tool by articulating lecturer's expectations from students	None	None
(Stevens, 2013)	School Students	To combine authentic and standardized measures of assessment to support student achievement and learning	None	None

(Hensel and Stanley, 2014)	University Students (Nursing Education)	To score student performance in a simulated authentic assessment	None	Inter-rater
(Morrissey, 2014)	School Students	To investigate how an authentic task could engage and motivate students	None	None
(Denisa, 2016)	Aerobics Instructor)	To emphasize authentic assessment within English for Specific Purposes (ESP) as a tool to evaluate the course	None	None
(Sambeka, Nahadi, and Sriyati, 2017)	School Students	To obtain the scientific information about increase of student's concept mastering in project-based learning that used authentic assessment	None	None
(Gulikers, 2006)	University Students (Social-work)	To measure the beliefs and perceptions of authentic assessment and its influence on student learning	Construct; Consequ-ential	None
(Cross, Greer, and Pearce, 1998)	School Students	To use authentic assessment as an intervention to aid students in demonstrating their comprehension of reading material	None	None
(Diaz, 1999)	School Students	To investigate if real-life application and self-evaluation improved student motivation	None	None
(Jensen and Klonicke, 1999)	School Students	To assess academic and social growth of special education students	None	None
(Craig and McCormick, 2002)	School Students	Improving student learning through authentic assessment	None	None
(Bullens, 2002)	School Students	To show an authentic picture of students' progress and abilities	None	None
(Barber, King, and Buchanan, 2015)	University Students (Education, Nursing, Health Care, Gaming, Business)	To examine the relationship between problem-based learning, authentic assessment and the role of community in fostering learning in digital contexts	None	None
(Paragae, Marhaeni, and Dantes, 2013)	School students	To analyse the authenticity of teacher-made assessment	None	None

APPENDIX 2: PAST APPROACHES TO VALIDITY AND RELIABILITY WHEN AUTHENTIC ASSESSMENT WAS IMPLEMENTED WITHOUT ASSESSMENT RUBRICS (SECOND LITERATURE REVIEW)

Author	Context of Implementation	Research Aims	Validity	Reliability
(Fall, 1996)	University Students (Public Relations)	To enhance students' levels of motivation and their use of higher order critical thinking skills through authentic assessment.	None	None
(Herrington and Herrington, 1998)	University Students (Mathematics)	To determine how students' respond to a model of authentic assessment.	None	None
(Saunders et al., 2001)	University Students (Education)	To determine whether a performance-based assessment system was related to the development of adult learners' cognitive skills.	None	None
(Palmer, 2004)	University Students (Engineering)	To determine whether authentic assessment can offer a more authentic representation of practice and assist in developing desirable skills.	None	None
(Balasubramanian, 2006)	School Students	To increase student achievement using an online communication and assessment tool.	None	None
(Scholtz, 2007)	University Students (Physiology)	To analyse the impact of authentic assessment on student performance in a technology-mediated constructivist classroom.	None	None
(Hallam et al., 2007)	School Students	To determine the effects of outcomes-driven authentic assessment on classroom quality.	None	None
(Tai and Yuen, 2007)	University Students (Multimedia)	To develop authentic assessment strategies in problem-based learning.	None	None
(Adeyemi, 2008)	School Students	To investigate the relationship between authentic and portfolio assessment based on the performance and attitude of the students towards the assessments.	None	None
(Ashford-Rowe, 2009)	Professional (Army)	To identify from the literature and codify into an applicable framework, the critical elements that would determine an assessment as being authentic.	None	None
(Gallagher and Shellshear, 2010)	University Students	To implement innovative and authentic assessment tasks	None	None

	(Teacher Education)	within pre-service teacher education courses.		
(Kovacs and Vacaretu, 2010)	Professional (Teachers)	To develop a guidebook for the authentic assessment of in-service teacher training.	None	None
(Fook and Sidhu, 2010)	University Students (Education)	To examine the impact of authentic assessment in addressing the mismatch between curriculum content and assessment practices.	None	None
(Davison, 2011)	University Students (Various)	To explore the relationships between authentic assessment and types of learner autonomy.	None	None
(Quartuch, 2011)	School Students	To determine civic engagement of students and teachers when authentic instruction was implemented.	None	None
(Vu, 2011)	University Students (Engineering, Ecology, etc.)	To experience the lived experience of authentic assessment practices in three university courses.	None	None
(Gao and Grisham-Brown, 2011)	School Students	To report accountability data on young children's language, literacy and pre-math competency.	Concurrent; Social	None
(Downing, 2012)	Online Education	To develop plurilingual competence through authentic and self-assessment.	None	None
(Koh et al., 2012)	School Students	To create thinking schools through authentic assessment.	None	None
(Baker, 2013)	Library Classroom	To define the attributes of authentic assessment and explain how this type of assessment can be used in a library classroom.	None	None
(Rourke and Coleman, 2011)	University Students (Fine Arts)	To examine the notion of authentic assessment and the role e-learning can play as a teaching and learning tool.	None	None
(Santos and Manuel, 2017)	University students (Pharmacy)	To describe and evaluate the design and implementation of an authentic assessment.	None	None
(Saunders, Saunders, and Batson, 2001)	University Students (Education)	To determine if authentic assessment was related to the development of adult learner's cognitive skills.	None	None

APPENDIX 3: INVESTIGATING EXISTENCE OF A CONCEPTUAL FRAMEWORK TO ADDRESS THE DIFFERENT ASPECTS OF VALIDITY AND RELIABILITY OF AUTHENTIC ASSESSMENT (SECOND LITERATURE REVIEW)

Author (Year)	Context of use	Validity Tested	Reliability tested	Conceptual framework used for validity and reliability testing
Moon et al., 2005	School students	Content	Inter-rater	None
Oh et al., 2005	University students	Content	Inter-rater; Internal consistency	None
Johnson, 2007	School students	Face; Content	Internal consistency	None
Olfos and Zulantay, 2007	School students	Criterion	Internal consistency	None
Jonsson, 2008	University students	Face; Construct	Inter-rater; Internal consistency	None
Diller and Phelps, 2008	University students	None	Internal consistency	None
Cassidy, 2009	University students	Construct	None	None
Koh and Luke, 2009	School students	None	Inter-rater	None
Taylor, 2011	School students	None	Inter-rater	None
Gao and Grisham-Brown, 2011	School students	Criterion	None	None
Fatonah et al., 2013	School students	Content	Inter-rater	None
Hensel and Stanley, 2014	University students	None	Inter-rater	None
Gulikers, 2006	University students	Content; Consequential	None	None
Lang II	University students	Content	None	None

APPENDIX 4: STUDENT PERCEPTION SURVEY

Please answer the following questions by placing a cross in only **one (1)** box:

Q1. Student ID No. _____

Q2. Work experience as a seafarer.

☐ Less than 3 years

☐ More than 3 years

Q3. English is my first language.

☐ Yes

☐ No

Q4. Highest level of education completed.

☐ Up to Year 12

☐ University education (bachelor's degree, master's degree, etc.)

Q5. This assessment was oriented to my future profession of a seafarer.

☐ Totally Disagree

☐ Disagree

☐ Neither Agree or Disagree

☐ Agree

☐ Totally Agree

Q6. This assessment was clearly directed to my professional requirements.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q7. This assessment prepared me for my future profession.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q8. The task of the assessment resembled the task of a real seafarer.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q9. The task of this assessment was an important part of the seafarer's profession.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q10. The tasks of this assessment differed from the tasks of a real seafarer.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q11. The context of in which I had to perform the assessment was fake.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q12. The context of in which I had to perform the assessment looked like a seafarer's workplace.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q13. The context of in which I had to perform the assessment looked just like the real world.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q14. The context of in which I had to perform the assessment was realistic.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q15. This way of assessing is an effective way of assessing professional skills.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q16. This way of assessing fits well with the seafarer's profession.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q17. The result (output) that I had to produce in this assessment is part of the seafarer's job.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q18. The output that was evaluated in this assessment is different from what is being evaluated in practice.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q19. The result that I had to produce in this assessment is something that a real seafarer also had to produce in practice.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q20. The criteria resembled the criteria that I have to meet in practice.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q21. The criteria that I had to meet in this assessment resembled the criteria used in practice.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q22. In this assessment I was evaluated on criteria important for the seafarer's profession.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q23. In this assessment I was evaluated on things that I never have to use in real professional practice.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q24. The criteria that I had to meet in this assessment were clear enough.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q25. Before I started the assessment, it was clear to me what was expected of me.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q26. It was hard to find out what was expected of me in this assessment.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q27. In this assessment, both knowledge and professional skills were important.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q28. The following requirements of the assessment helped me to improve my score:

a. Memorizing course material.

- ☐ Very much
- ☐ Quite a bit
- ☐ Some
- ☐ Very little

b. Applying facts, theories, or methods to real-world situations.

- ☐ Very much
- ☐ Quite a bit
- ☐ Some
- ☐ Very little

c. Analysing an idea, experience, or line of reasoning in depth by examining.

- ☐ Very much
- ☐ Quite a bit
- ☐ Some
- ☐ Very little

d. Evaluating a point of view, decision, or informed source.

- ☐ Very much
- ☐ Quite a bit
- ☐ Some
- ☐ Very little

e. Forming a new idea or understanding from various pieces of information.

- ☐ Very much
- ☐ Quite a bit
- ☐ Some
- ☐ Very little

Q29.

The feedback provided in this assessment helped me to identify the strengths and weaknesses in my learning.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

Q30.

a. This assessment provided more than one (1) opportunity to improve my score.

- ☐ Yes
- ☐ No

If 'No', the survey ends here for the respondents.

b. If 'Yes', the feedback provided on my first performance helped me to improve my assessment score in the next performance.

- ☐ Totally Disagree
- ☐ Disagree
- ☐ Neither Agree or Disagree
- ☐ Agree
- ☐ Totally Agree

This is the end of this survey. Many thanks for your time and contribution towards gaining a more informed understanding of student perceptions regarding authenticity in assessment.

APPENDIX 5: FACTORS EXTRACTED THROUGH EXPLORATORY FACTOR ANALYSIS (EFA) AND THEIR CONTRIBUTION TOWARDS EXPLAINING THE VARIANCE IN DATA

Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	11.536	38.454	38.454	11.195	37.317	37.317	8.949
2	2.330	7.767	46.221	2.003	6.677	43.995	6.295
3	2.009	6.696	52.917	1.543	5.142	49.137	8.709
4	1.707	5.691	58.608	1.303	4.342	53.479	4.631
5	1.333	4.444	63.052	.986	3.288	56.767	4.997
6	1.146	3.821	66.873	.780	2.601	59.368	3.062
7	1.028	3.426	70.299	.586	1.952	61.319	3.701
8	.903	3.009	73.308				
9	.861	2.870	76.178				
10	.773	2.577	78.754				
11	.693	2.311	81.065				
12	.640	2.133	83.198				
13	.604	2.013	85.212				
14	.540	1.800	87.012				
15	.464	1.547	88.559				
16	.426	1.421	89.980				
17	.409	1.364	91.343				
18	.344	1.147	92.491				
19	.333	1.111	93.602				
20	.312	1.039	94.641				
21	.289	.965	95.606				
22	.264	.880	96.486				
23	.217	.725	97.210				
24	.170	.566	97.776				
25	.143	.475	98.251				
26	.139	.463	98.715				
27	.129	.428	99.143				
28	.111	.369	99.512				
29	.100	.332	99.844				
30	.047	.156	100.000				

Extraction Method: Principal Axis Factoring.

a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.

Using the original survey questions (Question 5—30b) from Table 2, an initial factor analysis used a principle component analysis to reveal seven factors based on Eigenvalues greater than 1 that cumulatively explained 70.3% of the variance in data. The KMO and Bartlett's test revealed the values of 0.839 (Kaiser-Meyer-Olkin), 1844.494 (Bartletts's test of sphericity), 435 (df or degrees of freedom) and 0.000 (Significance value), which were considered acceptable for further analysis (Williams et al. 2010). A further exploratory factor analysis (EFA) with a direct-oblimin and promax rotation was also conducted. Hence, based on Eigenvalues greater than 1 (Sarkar et al. 2011) and using the scree plot generated, this research isolated seven factors.

APPENDIX 6: PATTERN MATRIX DETAILING FACTOR LOADING OF SURVEY QUESTIONS

	Factor						
	1	2	3	4	5	6	7
Q5	.910						
Q6	.844						
Q7	.639						
Q8	.829						
Q9	.465						
Q12				.819			
Q13				.668			
Q14				.705			
Q15			.760				
Q16			.558				
Q17	.630						
Q19			.711				
Q20			.500				
Q21			.519				
Q22	.568						
Q24					.558		
Q25					.817		
Q27			.546				
Q29							.843
Q30b							.499
Q28b		.514					
Q28c		.653					
Q28d		.774					
Q28e		.720					
Q10_R						.593	
Q11_R						.699	
Q23_R						.496	
Q26_R					.655		

Extraction Method: Principal Axis Factoring.
Rotation Method: Promax with Kaiser Normalization.
a. Rotation converged in 10 iterations.

The EFA used a promax rotation method and suppressed survey questions with coefficients less than 0.40 (Clark 2014) to obtain a pattern matrix that categorised the survey questions under the newly isolated seven factors. Due to suppression of questions with low coefficients, questions 18, 28a, and 30a were not included in the pattern matrix.

APPENDIX 7: MEAN AND STANDARD DEVIATION VALUES FOR STUDENT RESPONSES TO PERCEPTION SURVEY

Question	Mean	Standard deviation	Minimum	Maximum
Question 5	4.1	0.7	2	5
Question 6	4.0	0.8	2	5
Question 7	3.9	0.9	1	5
Question 8	4.0	0.7	2	5
Question 9	4.1	0.6	2	5
Question 10	3.5	0.9	2	5
Question 11	3.5	1.0	2	5
Question 12	3.9	0.7	2	5
Question 13	3.6	0.8	1	5
Question 14	3.9	0.7	2	5
Question 15	3.6	1.0	1	5
Question 16	3.7	1.0	1	5
Question 17	3.9	0.8	1	5
Question 18	3.1	1.0	2	5
Question 19	3.9	0.7	2	5
Question 20	3.9	0.8	1	5
Question 21	3.8	0.8	1	5
Question 22	4.0	0.6	2	5
Question 23	3.5	1.1	1	5
Question 24	3.7	1.0	1	5
Question 25	3.6	1.0	1	5
Question 26	3.2	1.1	1	5
Question 27	4.1	0.8	1	5
Question 28a	2.1	0.9	1	4
Question 28b	2.1	0.9	1	4
Question 28c	2.0	0.9	1	4
Question 28d	2	0.8	1	4
Question 28e	2.1	0.9	1	4
Question 29	3.5	1.0	1	5
Question 30a	1	0	1	1
Question 30b	3.9	0.8	2	5

STCW Competence: Use of leadership and managerial skills					
	High Distinction (>80%)	**Distinction (70-79%)**	**Credit (60-69%)**	**Pass (50-59%)**	**Fail (<50%)**
Task and Workload management	• Planned and co-ordinated task to share workload amongst all available crew members • Assigned required number of personnel towards task implementation accurately • Acknowledged and adhered to time and resource constraints without errors • Prioritized tasks according to the need of the situation correctly	• Planned and co-ordinated task to share workload amongst all available crew members • Assigned required number of personnel towards task implementation accurately • Acknowledged and adhered to time and resource constraints • Prioritized tasks according to the need of the situation with few errors	• Planned and co-ordinated task to share workload amongst all available crew members • Assigned required number of personnel towards task implementation accurately • Acknowledged and adhered to time and resource constraints with few errors • Prioritized tasks according to the need of the situation with few errors	• Planned and co-ordinated task to share workload amongst most of the available crew members • Assigned required number of personnel towards task implementation with few errors • Acknowledged and adhered to time and resource constraints with few errors • Prioritized tasks according to the need of the situation with few errors	• Failed to plan and co-ordinate task to share workload amongst crew members • Failed to assign required number of personnel towards task implementation • Failed to acknowledge and adhere to time and resource constraints • Failed to prioritize tasks according to the need of the situation
Resource management	• Allocated and assigned all available resources according to prioritization of need • Recognized all the barriers to effective communication and addressed them • Considered crew experiences in deciding course of action • Identified assertive leadership behaviours	• Allocated and assigned all available resources according to prioritization of need • Recognized all the barriers to effective communication and addressed them • Considered crew experiences in deciding course of action • Identified assertive leadership behaviours with few errors	• Allocated and assigned most of the resources according to prioritization of need • Recognized most of the barriers to effective communication and addressed them • Considered crew experiences in deciding course of action with few errors	• Allocated and assigned most of the resources according to prioritization of need • Recognized most of the barriers to effective communication and addressed them • Considered crew experiences in deciding course of action • Identified assertive leadership behaviours with few errors	• Failed to allocate and assign resources according to prioritization of need • Failed to recognize barriers to effective communication and address them • Failed to consider crew experiences in deciding course of action • Failed to identify assertive leadership behaviours • Failed to identify ways to motivate crew

	• Identified all possible ways to motivate crew • Obtained and maintained situation awareness accurately	• Identified all possible ways to motivate crew • Obtained accurate situation awareness but maintained it with few errors	• Identified assertive leadership behaviours with few errors • Identified few ways to motivate crew • Obtained and maintained situation awareness with few errors	• Overlooked ways to motivate crew • Obtained but did not maintain situation awareness at all	• Failed to obtain and maintain situation awareness
Decision-making	• Conducted situation analysis and risk assessment to identify all associated risks • Selected the right course of action based on the identified risks • Identified and generated optional course of action for all identified risks • Identified all necessary practices that evaluated outcome effectiveness for implemented course of action	• Conducted situation analysis and risk assessment to identify all associated risks • Selected the right course of action based on the identified risks • Identified and generated optional course of action for most of the identified risks • Implemented all necessary practices that evaluated outcome effectiveness for implemented course of action	• Conducted situation analysis and risk assessment to identify most of the associated risks • Selected the right course of action based on the identified risks • Identified and generated optional course of action for most of the identified risks • Identified most of the practices that evaluated outcome effectiveness	• Conducted situation analysis and risk assessment to identify most of the associated risks • Selected a course of action based on the identified risks with few errors • Identified and generated optional course of action for most of the identified risks with few errors • Failed to identify practices that evaluated outcome effectiveness	• Failed to conduct situation analysis and risk assessment • Failed to select the right course of action • Failed to identify and generate optional course of action • Failed to identify practices that evaluate outcome effectiveness
Comment:					

Social Science Ethics Officer
Private Bag 01 Hobart
Tasmania 7001 Australia
Tel: (03) 6226 2763
Fax: (03) 6226 7148
Katherine.Shaw@utas.edu.au

HUMAN RESEARCH ETHICS COMMITTEE (TASMANIA) NETWORK

UTAS

23 March 2017

Assoc Prof Benjamin Brooks
National Centre of Ports and Shipping
Australian Maritime College

Student Researcher: Samrat Ghosh

Sent via email

Dear Assoc Prof Brooks

Re: MINIMAL RISK ETHICS APPLICATION APPROVAL
Ethics Ref: **H0016321 - Authentic assessment in seafarer education: Correlation between student perceptions of assessment and their achievement**

We are pleased to advise that acting on a mandate from the Tasmania Social Sciences HREC, the Deputy Chair of the committee considered and approved the above project on 16 February 2017.

This approval constitutes ethical clearance by the Tasmania Social Sciences Human Research Ethics Committee. The decision and authority to commence the associated research may be dependent on factors beyond the remit of the ethics review process. For example, your research may need ethics clearance from other organisations or review by your research governance coordinator or Head of Department. It is your responsibility to find out if the approval of other bodies or authorities is required. It is recommended that the proposed research should not commence until you have satisfied these requirements.

Please note that this approval is for four years and is conditional upon receipt of an annual Progress Report. Ethics approval for this project will lapse if a Progress Report is not submitted.

The following conditions apply to this approval. Failure to abide by these conditions may result in suspension or discontinuation of approval.

1. It is the responsibility of the Chief Investigator to ensure that all investigators are aware of the terms of approval, to ensure the project is conducted as approved by the Ethics Committee, and to notify the Committee if any investigators are added to, or cease involvement with, the project.

2. Complaints: If any complaints are received or ethical issues arise during the course of the project, investigators should advise the Executive Officer of the Ethics Committee on 03 6226 7479 or human.ethics@utas.edu.au.

3. Incidents or adverse effects: Investigators should notify the Ethics Committee immediately of any serious or unexpected adverse effects on participants or unforeseen events affecting the ethical acceptability of the project.

4. Amendments to Project: Modifications to the project must not proceed until approval is obtained from the Ethics Committee. Please submit an Amendment Form (available on our website) to notify the Ethics Committee of the proposed modifications.

5. Annual Report: Continued approval for this project is dependent on the submission of a Progress Report by the anniversary date of your approval. You will be sent a courtesy reminder closer to this date. **Failure to submit a Progress Report will mean that ethics approval for this project will lapse.**

6. Final Report: A Final Report and a copy of any published material arising from the project, either in full or abstract, must be provided at the end of the project.

Yours sincerely

Katherine Shaw
Executive Officer
Tasmania Social Sciences HREC

Ethics Ref: H0016321
Title: Authentic assessment in seafarer education: Correlation between student perceptions of assessment and their achievement

This email is to confirm that the following amendment was approved by the Chair of the Tasmania Social Sciences Human Research Ethics Committee on 7/4/2017:

- Amendment for data collection in semester 2, 2017. The differing aspect between the assessments employed in Semester 1 to the assessment employed in semester 2 will be provided through an "authentic context".

All committees operating under the Human Research Ethics Committee (Tasmania) Network are registered and required to comply with the National Statement on Ethical Conduct in Human Research (NHMRC 2007, updated May 2015).

This email constitutes official approval. If your circumstances require a formal letter of amendment approval, please let us know.

Should you have any queries please do not hesitate to contact me.

Kind regards
Katherine

Katherine Shaw
Executive Officer, Social Sciences HREC
Office of Research Services | Research Division
University of Tasmania
Private Bag 1
Hobart TAS 7001
T +61 3 6226 2763
www.utas.edu.au/research

CRICOS 00586B

APPENDIX 10: CASE STUDIES USED FOR TRADITIONAL AND AUTHENTIC ASSESSMENT

The case studies in traditional assessment were presented devoid of context. The two case studies were implemented in a summative manner in traditional assessment, i.e. task 1 and task 2 were presented to the students at the same time.

Case studies (assessment tasks) in authentic assessment were presented in real-world contexts through a simulation of the scenarios by AMC staff. The two case studies were implemented formatively in authentic assessment, i.e. task 2 was implemented three weeks after task 1.

Case Study 1 (Task 1)

On sailing from the port of Sydney at night, you hear the emergency alarm being raised. It is announced that while securing the pilot ladder one of the seaman fell overboard on the starboard side of the vessel. Although a lifebuoy was thrown at sea, no one was sure if the seaman had managed to grab it. As one of the senior and experienced officers (Chief Officer/Second Engineer), you are leading the rescue team for an effective search and rescue operation. While the immediate action of saving life was imperative, the following conditions had to be considered:

- Other than you, the rescue team comprises of the Bosun, a junior engineer, and a deck officer;
- You find that the junior engineer is panicking and hesitating to accompany you on the rescue operation. This is creating a conflict between other rescue team members who are insisting him to perform his tasks as per his designated duties;
- The vessel did not carry a specialised rescue boat and the port lifeboat was the designated rescue boat.

1) Analyse the scenario and allocate tasks to the rescue team members (before and after the launching of the rescue boat). Include own responsibilities and duties as the team leader.
2) In this scenario, what factors will challenge you to conduct an effective search and rescue operation?
3) Develop strategies to overcome the identified challenges.
4) Identify the barriers to effective communication.

Case Study 2 (Task 2)

You are the Senior Officer (Master/Chief Officer or Chief Engineer/2nd Engineer) of a vessel with a compliment of 27 persons including you. The vessel is currently located in the Indian Ocean. Due to an uncontrollable fire on board at night, you had to abandon the vessel with 13 other crew members using the starboard side liferaft. The remaining crew members are expected to use the port side liferaft for the same purpose. However, you realise that some of the crew members panicked and jumped overboard before either of the liferafts were launched. Being in-command of the starboard side liferaft and responsible for the lives of the remaining crew members, you have the following considerations to decide on your next course of action:

- The port side liferaft has been launched with some of the crew members;
- There are no other vessels in sight to ask for assistance although the Master did send the 'Mayday' message using satellite communications; and
- Some of crew members, wearing life jackets, are now floating in water.

After the starboard side liferaft is launched at sea, describe your actions as the person-in command in the following aspects of leadership and managerial skills:
1) Prioritize your course of action (for the first hour after immediate abandon ship) as the leader of the starboard liferaft. In your answer, include the identified risks and the strategies to mitigate them.
2) In this scenario, what factors will challenge you to conduct an effective search and rescue of survivors at sea?
3) Suggest strategies to reduce panic.
4) Identify factors that may hinder detection of own liferafts at sea by other vessels in the vicinity; and determine strategies to address them.

REFERENCES

Abeywickrama, P 2012, 'Rethinking traditional assessment concepts in classroom-based assessment', *The CATESOL Journal,* vol. 23, no. 1, pp. 205-214.

Allianz Global Corporate & Specialty (AGCS) 2015, *Safety and shipping review 2015*, ALLIANZ, Munich, Germany.

AMSA 2011, *Australian Maritime Safety Authority: Corporate plan 2011-2016*, AMSA, Australian Government, Canberra.

AMSA 2013a, *Australian Maritime Safety Authority: Corporate plan 2013-18*, AMSA, Australian Government, Canberra.

AMSA 2013b, Certificates and endorsements: Immigration qualification assessment, Canberra, viewed 11 October 2013, http://www.amsa.gov.au/seafarers/int-qualifications/certs-endorsements-recognition/immigration-qualification-assessment/index.asp.

Andrade, H & Du, Y 2005, 'Student perspectives on rubric-referenced assessment', *Practical Assessment, Research & Evaluation*, vol. 10, no. 3, pp. 1-11.

Andrade, HG 2000, 'Using rubrics to promote thinking and learning', *Educational Leadership*, vol. 57, no. 5, pp. 13-19.

Aranda, S & Yates, P 2009, 'An overview of assessment', *The National Cancer Nursing education Project (EdCaN)*, Cancer Australia, Canberra.

Archbald, DA 1991, 'Authentic assessment: Principles, practices, and issues', *School Psychology Quarterly*, vol. 6, pp. 279-293.

Archbald, DA & Newmann, FM 1988, *Beyond standardized testing: Assessing authentic academic achievement in the secondary school*, National Association of Secondary Principals, Reston.

Ashford-Rowe, KH 2009, 'A heuristic framework for the determination of the critical elements in authentic assessment', PhD thesis, University of Wollongong.

Atkinson, RC & Shiffrin, RM 1968, 'Human memory: a proposed system and its control processes', in KW Spence & JT Spence (eds), *The Psychology of learning and motivation: Advances in research and theory*, Academic Press Psychology of Learning and Motivation, New York, vol. 2. pp. 89-195.

Australian Maritime College (AMC) 2011, *Report on the review of the national centre for ports & shipping: Deck officer courses,* AMC, Tasmania.

Azim, S & Khan, M 2012, 'Authentic assessment: An instructional tool to enhance students learning', *Academic Research International*, vol. 2, no. 3, pp. 314-320.

Bailey, KM 1998, *Learning about language assessment: dilemmas, decisions, and directions,* Heinle & Heinle, New York.

Bandura, A 1977, 'Self-efficacy: Toward a unifying theory of behavioural change', *Psychological Review*, vol. 84, no. 2, pp. 191-215.

Barber, W, King, S & Buchanan, S 2015, 'Problem based learning and authentic assessment in digital pedagogy: Embracing the role of collabrative communities', *The Electronic Journal of e-Learning*, vol. 13, no. 2, pp. 59-67.

Baylon, AM & Santos, VEMR 2011, 'The challenges in Philippine maritime education and training', *International Journal of Innovative Interdisciplinary Research*, no.1, pp. 34-43.

Bell, A & Bell, M 2003, 'Developing authentic assessment methods from a multiple intelligences perspective', Master of Arts Action Research Project, Saint Xavier University and SkyLight Professional Development Field-based master's Program.

Benner, P 2001, *From novice to expert: Excellence and power in clinical nursing practice*, Prentice-Hall, Inc, New Jersey.

Beyaztas, DI & Senemoglu, N 2015, 'Learning approaches of successful students and factors affecting their learning approaches', *Education and Science*, vol. 40, no. 179, pp. 193-216.

Bhardwaj, S 2009, 'Quality maritime education and training', in: V Loginovsky (ed.), *MET Trends in the XXI Century: Shipping industry and training institutions in the global environment – area of mutual interests and cooperation*, September 19-21, 2009, Admiral Makarov State Maritime Academy, Russia, pp. 29-32.

Biggs, J 1999, 'What the student does: teaching for enhanced learning', *Higher education research & development*, vol. 18, no. 1, pp. 57-75.

Biggs, J & Tang, C 2007, *Teaching for quality learning at university*, McGraw Hill, New York.

Biggs, J & Tang, C 2010, 'Applying constructive alignment to outcomes-based teaching and learning', *Training Material for "Quality Teaching for Learning in Higher Education" Workshop for Master Trainers*', February 23-25, 2010, Ministry of Higher Education, Kuala Lumpur.

Biggs, J & Tang, C 2011, *Teaching for quality learning at university: What the student does*, McGraw-Hill, England.

Black, P & Wiliam, D 1998a, 'Assessment and classroom learning', *Assessment in Education: Principles, Policy & Practice,* vol. 5, no.1, pp. 7-74.

Black, P & Wiliam, D 1998b, 'Inside the black box: Raising standards through classroom assessment', *Phi Delta Kappan*, vol. 92, no. 1, pp. 81-90.

Blackburn, JJ & Kelsey, KD 2013, 'Understanding authentic assessment in a secondary agricultural mechanics laboratory: An Instrumental Case Study', *Journal of Human Sciences and Extension*, vol. 1, no. 1, pp. 1-16.

Blondy, LC 2007, 'A correlational study between the critical thinking skills of nursing faculty and their perceived barriers to teaching critical thinking skills to nursing students', PhD thesis, Capella University.

Bloor, M & Sampson, H 2009, 'Regulatory enforcement of labour standards in an outsourcing globalized industry: the case of the shipping industry', *Work, Employment & Society*, vol. 23, no. 4, pp. 711-726.

BoarerPitchford, JK 2010, 'An examination of the assessment practices of community college instructors', PhD thesis, Caella University.

Bosco, AM & Ferns, S 2014, 'Embedding of authentic assessment in work-integrated learning curriculum', *Asia-Pacific Journal of Cooperative Education*, vol. 15, no. 4, pp. 281-290.

Boud, D & Falchikov, N 2006, 'Aligning assessment with long-term learning', *Assessment & Evaluation in Higher Education,* vol. 31, no. 4, pp. 399-413.

Boud, D & Walker, D 1998, 'Promoting reflection in professional courses: the challenge of context', *Studies in Higher Education*, vol. 23, no. 2, pp. 191-206.

Brady, D. 1997. 'Assessment and the curriculum', in C Cullingford (ed.), *Assessment versus evaluation*, Cassell, Great Britain.

Bransford, JD, Brown, AL & Cocking, RR 2000, 'Learning and transfer. How people learn: Brain, mind, experience, and school', National Academy Press, Washington, DC.

Brawley, N 2009, 'Authentic assessment vs. traditional assessment: A comparative study', Bachelor thesis, Coastal Carolina University.

Brindley, G 1998, 'Outcomes-based assessment and reporting in language learning programmes: a review of the issues', *Language Testing*, vol. 15, no. 1, pp. 45-85.

Bullens, D 2002, 'Authentic assessment: Change for the future', Master thesis, Saint Xavier University.

Burke, K 2011, *From standards to rubrics in six steps: Tools for assessing student learning*, Corwin, California.

Burton, K 2011, 'A framework for determining the authenticity of assessment tasks: Applied to an example in law', *Journal of Learning Design*, vol. 4, no. 2, pp. 20-28.

Burton, LJ & Mazerolle, SM 2011, 'Survey instrument validity Part I: Principles of survey instrument development and validation in athletic training education research', *Athletic Training Education Journal*, vol. 6, no. 1, pp. 27-35.

Cassidy, KE 2009, 'Using authentic intellectual assessment to determine level of instructional quality of teacher practice of new elementary school teachers based on teacher preparation route', PhD thesis, The George Washington University.

CEDEFOP 2010, *Learning outcomes approaches in VET curricula: A comparative analysis of nine European countries,* Publications Office of the European Union, Luxembourg.

Chance, BL 1997, 'Experiences with authentic assessment techniques in an introductory statistics course', *Journal of Statistics Education*, vol. 5, no. 3, pp. 1-15.

Chaudhry NG & Rasool, G 2012, 'A case-study on improving problem solving skills of undergraduate computer science students', *World Applied Sciences Journal*, vol. 20, no. 1, pp. 34-39.

Clanchy, J & Ballard, B 1995, 'Generic skills in the context of higher education', *Higher Education Research and Development*, vol. 14, no. 2, pp. 155-166.

Clark, RA 2014, 'Correlation study: The effect of student-teacher rapport on high school student performance rate', PhD thesis, Liberty University.

Coe, R 2002, 'It's the effect size, stupid: What effect size is and why it is important', *Annual Conference of the British Educational Research Association*, September 12-14, 2002, University of Exeter, England.

Cooper, BS & Gargan, A 2009, 'Rubrics in Education: Old Term, New Meanings', *Phi Delta Kappan*, vol. 91, no. 1, pp. 54-55.

Cooperstein, SE & Kocevar-Weidinger, E 2004, 'Beyond active learning: a constructivist approach to learning', *Reference Services Review*, vol. 32, no. 2, pp. 141-148.

Cousin, G 2006, 'An introduction to threshold concepts', *Planet*, vol. 17, pp. 4-5.

Cox, QN 2009, 'MET and Industry - Gaps to be bridged', in V Loginovsky (ed.), *MET trends in the XXI Century: Shipping industry and training institutions in the global environment – area of mutual interests and cooperation*, September 19-21, 2009, Admiral Makarov State Maritime Academy, Saint-Petersburg, pp. 171-181.

Cross, BJ, Greer, T & Pearce, M 1998, 'Improving student reading comprehension skills through the use of authentic assessment', Master research project, Saint Xavier University.

Cross, SJ 2007, 'Competence based learning and evaluation: Developments and non-developments in MET', *IFSMA 33rd Annual General Assembly*, May 24-25, 2007, Antwerp, Belgium.

Cumming, JJ & Maxwell, GS 1999. Contextualising authentic assessment. *Assessment in Education: Principles, Policies and Practices*, vol. 6, no. 2, pp. 177-194.

Curry, DH, Caplan, P & Knuppel, J 1994, 'Transfer of learning and adult learning (TOTAL)', *Journal of Continuing Social Work*, vol. 6, no. 1, pp. 8-14.

Darling-Hammond, L & Snyder, J 2000, 'Authentic assessment of teaching in context', *Teaching and Teacher Education*, vol. 16, no. 5, pp. 523-545.

Davis, MH, Amin, Z, Grande, JP, O'Neill, AE, Pawlina, W, Viggiano, TR & Zuberi, R 2007, 'Case studies in outcome-based education', *Medical Teacher*, vol. 29, no. 7, pp. 717-722.

Day, INZ, Blankenstein, FM, Westenberg, M & Admiral, W 2018, 'A review of the characteristics of intermediate assessment and their relationship with student grades', *Assessment & Evaluation in Higher Education,* vol. 43, no. 6, pp. 908-929.

Department of Education, Employment, and Workplace Relations (DEEWR) 2010, *Survey in the Maritime Industry - Oil and Gas Sector May 2010,* DEEWR, Canberra, Australia.

Denisa, K 2016, 'Authentic assessment in context of ESP', *International scientific & practical conference "world science"*, December 23-24, 2015, Issue 3(7), Abu Dhabi, UAE, pp. 24-28.

Dennison, RD, Rosselli, J & Dempsey, A 2015, *Evaluation beyond exams in nursing education: Designing assignments and evaluating with rubrics*, Springer Publishing Company, New York.

Department of Infrastructure and Transport 2012, *Department of infrastructure and transport submission to house of representatives standing committee on infrastructure and communications*, Australian Government, Canberra, Australia.

Diaz, C 1999, 'Authentic assessment: strategies for maximum performance in the dual-language classroom', Master thesis, Saint Xavier University.

Dikli, S 2003, 'Assessment at a distance: Traditional vs. alternative assessments', *The Turkish Online Journal of Educational Technology*, vol. 2, no. 3, pp. 13-19.

Diller, KR & Phelps, SF 2008, 'Learning outcomes, portfolios, and rubrics, Oh my! Authentic assessment of an information literacy program', *portal: Libraries and the Academy*, vol. 8, no. 1, pp. 75-89.

Donovan, MS, Bransford, JD & Pellegrino, JW 1999, *How people learn: Bridging research and practice*, National Academy Press, Washington, DC.

Driscoll, A & Wood, S 2007, *Developing Outcomes-based assessment for learner cantered education: A Faculty introduction*, Stylus Publishing, Virginia.

Drost, EA 2011, 'Validity and reliability in social science research', *Education Research and Perspectives*, vol. 38, no. 1, pp. 105-123.

Drown, D, Mercer, R, Jeffery, G & Cross, S 2010, 'Disparate measures in examinations for STCW certificates of competency: The use and effectiveness of multiple-choice questions', 4*th International Conference on Maritime Human Resource Solutions*, September 28-30, 2010, Canada.

Dunn, L, Parry, S & Morgan, C 2002, 'Seeking quality in criterion referenced assessment', *Learning Communities and Assessment Cultures Conference*, August 28-30, 2002, EARLI Special Interest Group on Assessment and Evaluation, University of Northumbria.

Einbender, L & Wood, D 1995, *An authentic journey: Teachers' emergent understanding about authentic assessment and practice*, National Centre for Restructuring Education, Schools, and Teaching (NCREST), New York.

Emad, G & Roth, WM 2007, 'Evaluating the competencies of seafarers: Challenges in current practice', in T Pelton, G Reis, & K Moore (eds.), *Proceedings of the University of Victoria Faculty of Education Research Conference - Connections'07*, 2007, University of Victoria, Victoria, BC, Canada, pp. 71-76.

Emad, G & Roth, WM 2008, 'Contradictions in the practices of training for and assessment of competency', *Journal of Education and Training*, vol. 50, no. 3, pp. 260-272.

Emery, DE 2001, 'Authentic assessment in high school science: A classroom perspective', in DP Shepardson (ed.), *Assessment in Science A Guide to Professional Development and Classroom Practice*, Springer-Science+Business Media Dordrecht, Indiana, pp. 227-248.

Endedijk, MD & Vermunt, JD 2013, 'Relations between student teachers' learning patterns and their concrete learning activities', *Studies in Educational Evaluation*, vol. 39**,** no. 1, pp. 56-65.

Ernstsen, J & Nazir, S, 2018, 'Human error in pilotage operations', TransNav: International Journal on Marine Navigation and Safety of Sea Transportation, vol. 12, no. 1, pp. 49-56.

Ertmer, PA & Newby, TJ 1993, 'Behaviourism, cognitivism, constructivism: Comparing critical features from an instructional design perspective', *Performance Improvement Quarterly*, vol. 6, no. 4, pp. 50-72.

Fatonah, S, Suyata, P & Prasetyo, ZK 2013, 'Developing an authentic assessment model in elementary school science teaching', *Journal of Education and Practice*, vol. 4, no. 13, pp. 50-61.

Finch, AE 2002, 'Authentic assessment: Implications for EFL performance testing in Korea', *Secondary Education Research*, vol. 49, pp. 89-122.

Findlay, LAE 2013, 'A Qualitative Investigation into student and teacher perceptions of motivation and engagement in the secondary mathematics classroom', Bachelor thesis, Avondale College of Higher Education.

Flin, RH, O'Connor, P & Crichton, M 2008, *Safety at the sharp end: a guide to non-technical skills*, Ashgate Publishing, Ltd., London.

Fook, CY & Sidhu, GK 2010, 'Authentic assessment and pedagogical strategies in higher education', *Journal of Social Sciences*, vol. 6, no. 2, pp. 153-161.

Forneris, SG & Peden-McAlpine, CJ 2006, 'Contextual learning: A reflective learning intervention for nursing education', *International Journal of Nursing Education Scholarship*, vol. 3, no. 1, pp. 1-18.

Fredricks, JA & McColskey, W 2012, 'The measurement of student engagement: A comparative analysis of various methods and student self-report instruments', in SL Christenson (ed.), *Handbook of Research on Student Engagement*, Springer Science+Business Media, New York, pp. 763-782.

Frey, BB, Schmitt, VL & Allen, JP, 2012, 'Defining authentic classroom assessment', *Practical Assessment, Research and Evaluation*, vol. 17, no. 2, pp. 1-18.

Gallagher, SA, Stepien, WJ & Rosenthal, H 1992, 'The effects of problem-based learning on problem solving', *Gifted Child Quarterly,* vol. 36**,** pp. 195-200.

Gao, X & Grisham-Brown, J 2011, 'The use of authentic assessment to report accountability data on young children's language, literacy and pre-math competency', *International Education Studies*, vol. 4, no. 2, pp. 41-53.

Ghosh, S, Bowles, M, Ranmuthugala, D & Brooks, B 2014a, 'On a lookout beyond STCW: Seeking standards and context for the authentic assessment of seafarers', in D Ranmuthugala, & B Lewarn (eds.), *IAMU AGA 15 Looking Ahead: Innovation in maritime education, training and research*, October 27-30, 2014a, Australian Maritime College, Launceston, Tasmania, pp. 77-86.

Ghosh, S, Bowles, M, Ranmuthugala, D & Brooks, B 2014b, 'Reviewing seafarer assessment methods to determine the need for authentic assessment', *Australian Journal of Maritime & Ocean Affairs*, vol. 6, no. 1, pp. 49-63.

Ghosh, S, Bowles, M, Ranmuthugala, D & Brooks, B 2015, 'Using authentic assessment to enhance seafarer student engagement and their ability to transfer learning', in B Svilicic & B Pritchard (eds.), *IAMU AGA 2015 International Association of Maritime Universities 16th IAMU Annual General Assembly*, October 7-10, 2015, University of Rijeka, Faculty of Maritime Studies, Opatija, Croatia, pp. 93-98.

Ghosh, S, Bowles, M, Ranmuthugala, D & Brooks, B 2016, 'Authentic assessment in seafarer education: Using literature review to investigate its' validity and reliability through rubrics', *WMU Journal of Maritime Affairs,* vol. 15, no. 2, pp. 317-330.

Ghosh, S, Bowles, M, Ranmuthugala, D & Brooks, B 2017, 'Improving the validity and reliability of authentic assessment in seafarer education and training: A conceptual and practical framework to enhance resulting assessment outcomes', *WMU Journal of Maritime Affairs,* vol. 16, no. 3, pp. 455-472.

Ghosh, S, Brooks, B, Ranmuthugala, D & Bowles, M To be published, 'Authentic versus traditional assessment: an empirical study investigating the difference in seafarer students' academic achievement'.

Gonczi, A, Hager, P & Athanasou, J 1993, *The development of competency-based assessment strategies for the profession,* Department of Employment, Education, and Training, Australian Government Publishing Service, Australia.

Goodwin, LD & Leech, NL 2006, 'Understanding correlation: factors that affect the size of r', *The Journal of Experimental Education*, vol. 74, no. 3, pp. 251-266.

Graziano, AM & Raulin, ML 2000, *Research methods: A process of inquiry,* A Pearson Education Company, United States of America.

Gulikers, JTM 2006, 'Authenticity is in the eye of the beholder: Beliefs and perceptions of authentic assessment and the influence on student learning', PhD thesis, Open Universiteit Nederland.

Gulikers, JTM, Bastiaens, TJ & Kirschner, PA 2004a 'A five-dimensional framework for authentic assessment', *ETR&D-Educational Technology Research and Development*, vol. 52, no. 3, pp. 67-86.

Gulikers, JTM, Bastiaens, TJ & Kirschner, PA 2004b, 'Perceptions of authentic assessment: Five dimensions of authenticity', *Second Biannual Joint Northumbria/EARLI SIG assessment conference*, University of Bergen, Norway, pp. 67-86.

Gulikers, JTM, Bastiaens, TJ & Kirschner, PA 2006, 'Authentic assessment, student and teacher perceptions: the practical value of the five-dimensional framework', *Journal of Vocational Education and Training*, vol. 58, no. 3, pp. 337-357.

Gulikers, JTM, Bastiaens, TJ, Kirschner, PA & Kester, L 2008, 'Authenticity is in the eye of the beholder: student and teacher perceptions of assessment authenticity', *Journal of Vocational Education and Training*, vol. 60, no.4, pp. 401-412.

Hanzu-Pazara, R & Arsenie, P 2007, 'Shipping companies' policy to improve the seafarer's competency', in D Zhukov (ed.), *8th IAMU Annual General Assembly*, September 17-19, 2007, Odesa, National Maritime Academy Odesa, Ukraine, pp. 313-320.

Hart, C, Hammer, S, Collins, P & Chardon, T 2011, 'The real deal: using authentic assessment to promote student engagement in the first and second years of a regional law program', *Legal Education Review*, vol. 21, no. 1, pp. 97-121.

Hattie, JAC & Clarke, S 2019, *Visible learning feedback*, Routledge, London.

Hattie, JAC 2009, *Visible learning: A synthesis of over 800 meta analyses relating to achievement,* Routledge, London.

Hattie, JAC & Timperley, H 2007, 'The power of feedback', *Review of Educational Research,* vol. 77**,** pp. 81-112.

Hensel, D & Stanley, L 2014, 'Group simulation for "authentic" assessment in a maternal-child lecture course', *Journal of the Scholarship of Teaching and Learning*, vol. 14, no.2, pp. 61-70.

Herrington, J & Herrington, A 1998, 'Authentic assessment and multimedia: How university students respond to a model of authentic assessment', *Higher Education Research and Development Society of Australasia*, vol. 17, no. 3, pp. 305-322.

Herrington, J & Herrington, A 2006, 'Authentic conditions for authentic assessment: Aligning task and assessment', *Critical Visions, Proceedings of the 29th HERDSA Annual Conference*, July 10-12, 2006, Higher Education Research and Development Society of Australasia, Inc., Western Australia, Australia, pp. 146-151.

Herrington, J, Reeves, TC & Oliver, R 2010, *A guide to authentic e-learning*, Taylor & Francis, United Kingdom.

Herrington, JA 1997, 'Authentic learning in interactive multimedia environments', PhD thesis, Edith Cowan University.

Hetherington, C, Flin, R & Mearns, K 2006, 'Safety in shipping: The human element', *Journal of Safety Research*, vol. 37, no.4, pp. 401-411.

Hoepfl, M 2000, 'Large-Scale Authentic assessment', in RL Custer (ed.), *Using authentic assessment in vocational education,* Centre on Education and Training for Employment, Columbus.

Huang, V 2017, 'An Australian study comparing the use of multiple-choice questionnaires with assignments as interim, summative law school assessment', *Assessment & Evaluation in Higher Education,* vol. 42, no. 4, pp. 580-595.

International Maritime Organization (IMO) 2002, *International safety management (ISM) code and revised guidelines on implementation of the ISM Code*, International Maritime Organization, London.

International Maritime Organization (IMO) 2011, *International convention on standards of training, certification and watchkeeping for seafarers,* International Maritime Organization, London.

International Maritime Organization (IMO) 2013a, *Development of an e-navigation strategy implementation plan*, Report of the Correspondence Group on e-navigation to Nav 59, Rio.

International Maritime Organization (IMO) 2013b, International convention on standards of training, certification and watchkeeping for Seafarers (STCW), Viewed 19 Aug 2013, http://www.imo.org/About/Conventions/ListOfConventions/Pages/International-Convention-on-Standards-of-Training,-Certification-and-Watchkeeping-for-Seafarers-%28STCW%29.aspx.

James, LT & Casidy, R 2016, 'Authentic assessment in business education: its effects on student satisfaction and promoting behaviour', *Studies in Higher Education*, vol. 43, no. 3, pp. 401-415.

Jensen, WM & Klonicke, AS 1999, 'Assessing social and academic skills of special education students using authentic assessment', Master thesis, Saint Xavier University.

Johnson, YL 2007, 'The Efficacy of authentic assessment versus pencil and paper testing in evaluating student achievement in a basic technology course', PhD thesis, Walden University.

Jonsson, A 2008, 'Educative assessment for/of teacher competency. A study of assessment and learning in the "Interactive examination" for student teachers', PhD Thesis, Malmo University.

Jonsson, A & Svingby, G 2007, 'The use of scoring rubrics: Reliability, validity and educational consequences', *Educational Research Review*, vol. 2, no. 2, pp. 130-144.

Kainth, S 2019, *Victoria pushes for tougher English requirements for international students in Australia*, SBS, viewed 29 May 2019, https://www.sbs.com.au/yourlanguage/punjabi/en/article/2019/01/24/victoria-pushes-tougher-english-requirements-international-students-australia.

Kean, K, Matthews, A, Meadows, J & Stone, H 2011, 'Testing the waters: Are marine evaluation techniques identifying competent officers?', *Nautical Science Technological Report*, The Company of Master Mariners, Fisheries and Marine Institute of Memorial University of Newfoundland, Canada.

Kearney, S 2012, 'Improving engagement: The use of 'Authentic self -and peer -assessment for learning' to enhance the student learning experience', *Assessment & Evaluation in Higher Education,* vol. 38**,** no. 7, pp. 875-891.

Kim, TE & Nazir, S 2016, 'Exploring marine accident causation: A case study', *Occupational Safety and Hygiene*, vol. IV, pp. 369-373.

Kim, TE, Nazir, S & Øvergård, KI 2016, 'A STAMP-based causal analysis of the Korean Sewol ferry accident', Safety science, vol. 83, pp. 93-101.

Kluge, A 2008, 'What you train is what you get? Task requirements and training methods in complex problem-solving', *Computers in human behaviour*, vol. 24, no. 2, pp. 284-308.

Kluge, A, Nazir, S & Manca D 2014, 'Advanced applications in process control and training needs of field and control room operators', *IIE Transactions on Occupational Ergonomics and Human Factors*, vol. 2, no. 3-4, pp. 121-136.

Koh, K & Luke, A 2009, 'Authentic and conventional assessment in Singapore schools: An empirical study of teacher assignments and student work', *Assessment in Education: Principles, Policy and Practice*, vol. 16, no. 3, pp. 291-318.

Koh, KH, Tan, C & Ng, PT 2012, 'Creating thinking schools through authentic assessment: the case in Singapore', *Educational Assessment, Evaluation and Accountability,* vol. 24, no. 2, pp. 135-149.

Lang II, TR 2012, 'An examination of the relationship between elementary education teacher candidates' authentic assessments and performance on the professional education subtests on the Florida teacher certification exam (FTCE)', Educational Specialist Graduate Theses and Dissertations, University of South Florida.

Law, B & Eckes, M 1995, *Assessment and ESL: On the yellow big road to the withered of Oz,* Peguis, Manitoba.

Leberman, SI 1999, 'The transfer of learning from the classroom to the workplace: A New Zealand case study', PhD thesis, Victoria University.

Leca Da Veiga, JPFA 2001, 'A study of the implementation of STCW 95 in the context of a safety culture in shipping', PhD thesis, University of Wales.

Lenz, TL, Monaghan, MS, Wilson, AF, Tilleman, JA, Jones, RM & Hayes, MM 2006, 'Using performance-based assessments to evaluate parity between a campus and distance education pathway', American Journal of Pharmaceutical Education, vol. 70, no. 4, article 90, pp. 1-6.

Lewarn, B 2002, 'Seafarer training–does the system defeat competence?', *Proceedings of International Association of Maritime Universities 3rd General Assembly*, September 23-26, 2002, Rockport, Maine.

Linn, RL, Baker, EL & Dunbar, SB 1991, 'Complex, performance-based assessment: Expectations and validation criteria', *Educational Researcher*, vol. 20, no. 8, pp. 15-21.

Lister, R 2006, 'Driving learning via criterion-referenced assessment using Bloom's taxonomy', *Assessment in Science Teaching and Learning Symposium*, September 28, 2006, The University of Sydney, UniServe SCIENCE, Sydney, Australia, pp. 80-88.

Listewnik, J 2009, 'The MET institutions in the forefront to realize the IMO motto "Go to sea campaign", in V Loginovsky (ed.), *MET Trends in the XXI Century: Shipping Industry and Training Institutions in the global environment – area of mutual interests and cooperation*, September 19 – 21, 2009, Admiral Makarov State Maritime Academy, Saint-Petersburg, Russia, pp. 33-42.

Lodico, MG, Spaulding, DT & Voegtle, KH 2010, *Methods in educational research: From theory to practice,* Jossey-Bass, San Francisco.

Lovorn, MG & Rezaei, AR 2011, 'Assessing the assessment: Rubrics training for pre-service and new In-service teachers', *Practical Assessment, Research & Evaluation*, vol. 16, no. 16, pp. 1-18.

Lukacs, PM, Burnham, KP & Anderson, DR 2010, 'Model selection bias and Freedman's paradox', *Ann Inst Stat Math*, vol. 62**,** pp. 117-125.

Lund, J 1997, 'Authentic assessment: Its development & applications', *Journal of Physical Education, Recreation & Dance*, vol. 68, no. 7, pp. 25-28.

Lynch, R 2003, 'Authentic, performance-based assessment in ESL/EFL reading instruction', *Asian EFL Journal*, vol. 49, pp. 89-122.

Mallet, DG 2006, 'Authentic assessment for advanced undergraduate students', *ANZIAM,* vol. 47**,** pp. C524-C540.

Maltby, A & Mackie, S 2009, 'Virtual learning environments - help or hindrance for the 'disengaged' student?', *ALT-J,* vol. 17, no. 1, pp. 49-62.

Maringa, MTH 2015, 'Assessment of quality of training and education of seafarers in South Africa and Ghana', Master thesis, World Maritime University.

Marzano, RJ, Pickering, D & McTighe, J 1993, *Assessing student outcomes:* Performance, Association for Supervision and Curriculum Development, Virginia.

Mbawo, E 1995, 'Strategies for enhancing transfer of training in the workplace', *Training and Management Development Methods*, vol. 9, no. 5, pp. 729-744.

McAlister, B 2001, 'The authenticity of authentic assessment: What the research says... or doesn't say', in RL Custer (ed.), *Using authentic assessment in vocational education,* Centre on Education and Training for Employment College of Education, Columbus, pp. 19-30.

McAlpine, M 2002, *Principles of assessment,* in CAA Centre (ed.), Blue paper Number 1, University of Glasgow, United Kingdom.

McCarthy, G 2013, 'Authentic assessment - key to learning', in E Doyle, P Buckley & C Carroll (eds.), *Innovative business school teaching - engaging the millennial generation*, Routledge, United Kingdom, pp. 81-92.

McLaughlin, H 2015, 'Seafarers in the spotlight', *Maritime Policy & Management*, vol. 42, no. 2, pp. 95-96.

McLeod, S 2013, What is reliability? Viewed 24 July 2016, http://www.simplypsychology.org/reliability.html.

McMullen, C & Braithwaite, I 2005, 'Authentic' assessment: Workplace based formal learning setting the stage for ongoing informal learning', *ANZMAC 2005 Conference: Marketing Education*, December 5-7, 2005, Fremantle, Australia, pp. 130-136.

Merriam, SB & Leahy, B 2005, 'Learning transfer: A review of the research in adult education and training', *PAACE Journal of Lifelong Learning*, vol. 14, no. 1, pp. 1-24.

Messick, S 1994, 'The interplay of evidence and consequences in the validation of performance assessments', *Educational Researcher*, vol. 23, no. 2, pp. 13-23.

Messick, S 1995, 'Validity of psychological assessment: Validation of references from person's responses and performances as scientific inquiry into score meaning', *American Psychologist*, vol. 50, no. 9, pp. 741-749.

Messick, S 1996, 'Validity and washback in language testing', *Language Testing*, vol. 13, no. 3, pp. 241-256.

Mestre, J 2002, 'Transfer of learning: Issues and research agenda', *Report of a Workshop held at the National Science Foundation*, University of Massachusetts-Amherst, Virginia.

Meyer, CA 1992, 'What's the difference between "authentic" and "performance" assessment?', *Educational leadership*, vol. 49, no. 8, pp. 39-40.

Mitchell, GW 2008, 'Essential soft skills for success in the twenty-first century workforce as perceived by Alabama business/marketing educators', PhD thesis, Auburn University.

Montgomery, K 2002, 'Authentic tasks and rubrics: Going beyond traditional assessments in college teaching', *College teaching*, vol. 50, no. 1, pp. 34-40.

Moon, TR, Brighton, CM, Callahan, CM & Robinson, A 2005, 'Development of authentic assessments for the middle school classroom', *The Journal of Secondary Gifted Education*, vol. 16, no. 2/3, pp. 119-133.

Moore, JL 2012, 'Designing for transfer: A threshold concept', *The Journal of Faculty Development*, vol. 26, no. 3, pp. 19-24.

Moore, TJ & Asay, SM 2013, *Values, attitudes and behaviours: Understanding family choices. Family Resource Management*. SAGE Publications, United States.

Morrison, WGS 1997, *Competent crews = safer ships; An aid to understanding STCW 95*, WMU Publications, Malmo, Sweden.

Morrissey, PE 2014, 'Investigating how an authentic task can promote student engagement when learning about Australian history', PhD thesis, University of Wollongong.

Moskal, BM 2000, 'Scoring rubrics: what, when and how?', *Practical Assessment, Research & Evaluation*, vol. 7, no. 3, pp. 1-5.

Moskal, BM & Leydens, JA 2000, 'Scoring rubric development: Validity and reliability', *Practical Assessment, Research & Evaluation*, vol. 7, no. 10, pp. 71-81.

Mueller, J 2005, 'The authentic assessment toolbox: Enhancing student learning through online faculty development', *Journal of Online Learning and Teaching*, vol. 1, no.1.

Mueller, J 2006, Authentic assessment toolbox, viewed 6 February 2019. http://jonathan.mueller.faculty.noctrl.edu/toolbox/whatisit.htm#looklike.

Munns, G & Woodward, H 2006, 'Student engagement and student self-assessment: the REAL framework', *Assessment in Education*, vol. 13, no. 2, pp. 193-213.

Nazir, S, Jungefeldt, S & Sharma, A 2019, 'Maritime simulation training across Europe: a comparative study', WMU Journal of Maritime Affairs, vol. 18, no. 1, pp. 197-224.

Nazir, S, Sorensen, LJ, Øvergård, KI & Manca, D 2015, 'Impact of training methods on Distributed Situation Awareness of industrial operators', *Safety science*, vol. 73, pp. 136-145.

Neely, P & Tucker, J 2012, 'Using business simulations as authentic assessment tools', *American Journal of Business Education*, vol. 5, no. 4, pp. 449-456.

Neill, J 2008, *Writing up a factor analysis,* University of Canberra, Canberra.

Neo, M, Neo, KT-K & Tan, HY-J 2012, 'Applying authentic learning strategies in a multimedia and web learning environment (MWLE) Malaysian students' perspective', *The Turkish Online Journal of Educational Technology*, vol. 11, no. 3, pp. 50-60.

Newmann, FM, Marks, HM & Gamoran, A 1996, 'Authentic pedagogy and student performance', *American Journal of Education*, vol. 104, no. 4, pp. 280-312.

Nusche, D 2008, *Assessment of learning outcomes in higher education: A comparative review of selected practices*, OECD Education Working Paper No.15, OECD Directorate of Education, France.

Oh, DM, Kim, JM, Garcia, RE & Krilowicz, BL 2005, 'Valid and reliable authentic assessment of culminating student performance in the biomedical sciences', *Advances in Physiology Education*, vol. 29, no. 2, pp. 83-93.

Olfos, R & Zulantay, H 2007, 'Reliability and validity of authentic assessment in a web-based course', *Educational Technology & Society*, vol. 10, no. 4, pp. 156-173.

Osborne, R, Dunne, E & Farrand, P 2013, 'Integrating technologies into ''authentic'' assessment design: an affordances approach', *Research in Learning Technology*, vol. 21, no. 1, pp. 1-18.

Ozimek, DJ 2000, 'Student learning, retention, and transfer from trigonometry to physics', Master thesis, Kansas State University.

Pallis, AA & Ng, AKY 2011, 'Pursuing maritime education: an empirical study of students' profiles, motivations and expectations', *Maritime Policy & Management*, vol. 38, no. 4, pp. 369-393.

Palm, T 2008, 'Performance assessment and authentic assessment: A conceptual analysis of the literature', *Practical Assessment, Research and Evaluation*, vol. 13, no. 4, pp. 1-11.

Palomba, CA & Banta, TW 1999, *Assessment essentials: Planning, implementing, and improving assessment in higher education*, Jossey-Bass, San Francisco.

Pecota, SR & Buckley, JJ 2009, 'Training paradigm assisted accidents: Are we setting our students up For failure?', in V LOGINOVSKY (ed.), *MET Trends in the XXI Century: Shipping industry and training institutions in the global environment – area of mutual interests and cooperation*, September 19-21, 2009, Admiral Makarov State Maritime Academy, Russia, pp. 192-204.

Prasad, R 2011, 'The role of collaborative learning in promoting safety in shipboard machinery spaces', PhD thesis, World Maritime University.

Prasad, R, Nakazawa, T & Baldauf, M 2010, 'Professional Development of Shipboard Engineers and the Role of Collaborative Learning', *International Association of Maritime Universities AGA11*, October 15-18, 2010, Korea Maritime University, Busan, Korea, pp. 165-174.

Quartuch, MJ 2011, 'Is Authentic Enough? Authentic assessment and civic engagement', Master thesis, Moravian College.

Reason, J 2000, 'Human error: models and management', *BMJ*, vol. 320, no. 7237, pp. 768-770.

Reason, J 1995, 'Understanding adverse events: human factors', *BMJ Quality & Safety*, vol. 4, no. 2, pp. 80-89.

Reason, J, Hollnagel, E & Paries J 2008, 'Revisiting the Swiss cheese model of accidents', *Journal of Clinical Engineering*, vol. 27, no. 4, pp. 110-115.

Reddy, MY 2007, 'Rubrics and the enhancement of student learning', *Educate~*, vol. 7, no. 1, pp. 3-17.

Reeves, TC & Okey, JR 1996, *Alternative assessment for constructivist learning environments*, Educational Technology Publications, Englewood Cliffs, New Jersey.

Rezaei, AR & Lovorn, M 2010, 'Reliability and validity of rubrics for assessment through writing', *Assessing Writing*, vol. 15, no. 1, pp. 18-39.

Rhodes, TL & Finley, A 2013, *Using the VALUE rubrics for improvement of learning and authentic assessment*, Association of American Colleges and Universities, United States of America.

Richards Perry, GD 2011, 'Student perceptions of engagement in schools: A Deweyan analysis of authenticity in high school classrooms', PhD thesis, Georgia State University.

Robson, C 2011, *Real world research,* John Wiley & Sons, Chichester.

Robson, CS 2007, 'Toward an international rubric: A compilation of STCW competency assessment methodologies', in D Zhukov, (ed.), *8th IAMU Annual General Assembly*, Odesa National Maritime Academy, Odesa, Ukraine, pp. 247-258.

Rourke, AJ & Coleman, KS 2011, 'Authentic assessment in e-learning: Reflective and collaborative writing in the arts', in G WILLIAMS, P. STATHAM, N BROWN, & B CLELAND (eds.), *Changing Demands, Changing Directions*, December 4-7, 2011, Proceedings ascilite Hobart, Hobart, Australia, pp. 1089-1095.

Rutherford, PD 1995, *Competency based assessment: A guide to implementation*, Pitman Publishing Asia Pacific, Australia.

Rychen, D 2004, 'Key competencies for all: an overarching conceptual frame of reference', in DS Rychen, & A Tiana (eds.), *Developing key competencies in education: some lessons from international and national experience*. UNESCO International Bureau of Education, Paris, pp. 5-34.

Sadjadi, J & Perkins, SJ 2010, 'The human element in international seafaring', University of Bedfordshire Business School, United Kingdom.

Sadler, PM & Good, E 2006, 'The impact of self- and peer-grading on student learning', *Educational Assessment,* vol. 11, no. 1, pp. 1-31.

Sadler, R 2005, 'Interpretations of criteria-based assessment and grading in higher education', *Assessment & Evaluation in Higher Education*, vol. 30, no. 2, pp. 175-194.

Sambeka, Y, Nahadi & Sriyati, S 2017, 'Implementation of authentic assessment in the project-based learning to improve student's concept mastering', in T Hidayat, & E Al (eds.), *Mathematics, Science, and Computer Science Education (MSCEIS 2016)*, October 15, 2016, American Institute of Physics, Bandung, Indonesia, pp. 060012-1 to 060012-6.

Sampson, D & Fytros, D 2008, 'Competence models in technology-enhanced competence-based learning', in Adelsberger, H Heimo, Kinshuk, Pawlowski, J Martin (eds.), *Handbook on information technologies for education and training,* Springer, Heidelberg, pp. 155-177.

Sampson, H 2004, 'Romantic rhetoric, revisionist reality: the effectiveness of regulation in maritime education and training', *Journal of Vocational Education and Training*, vol. 56, no. 2, pp. 245-268.

Sampson, H, Gekara, V & Bloor, M 2011, 'Water-tight or sinking? A consideration of the standards of the contemporary assessment practices underpinning seafarer licence examinations and their implications for employers', *Maritime Policy & Management: The flagship journal of international shipping and port research*, vol. 38, no. 1, pp. 81-92.

Santos, S & Manuel, J 2017, 'Design, implementation and evaluation of an authentic assessment experience in a pharmacy course: are students getting it?', in JDI Soria, MCV Vela, EDL Poza & D Blazquez (eds.), *3rd International Conference on Higher Education Advances, HEAd'17, June 21-23, 2017,* Universitat Polit`ecnica de Val`encia, Valencia, Spain, pp. 574-583.

Sarkar, E, Keskin, S & Unver, H 2011, 'Using of factor analysis scores in multiple linear regression model for prediction of kernel weight in ankara walnuts', *The Journal of Animal & Plant Sciences,* vol. 21, no. 2, pp. 182-185.

Sator, A 2000, 'An Exploration of Transfer of Learning Opportunities in an Online Co-Operative Prepatory Curriculum', Master of Arts thesis, Simon Fraser University.

Saunders, NG, Saunders, GA & Batson, T 2001, 'Assessment and the adult learner: Does authentic assessment influence learning?', *Annual Meeting of the Mid-Western Educational Research Association,* October 24-27, 2001, Chicago, IL, pp. 1-22.

Schell, JW 2000, 'Think about authentic learning and then authentic assessment', in RL Custer (ed.), *Using authentic assessment in vocational education,* Centre on Education and Training for Employment, Columbus, pp. 7-18.

Schneider, RM, Krajcik, J, Marx, R & Soloway, E 2002, 'Performance of students in project-based science classrooms on a national measure of science achievement', *Journal of Research in Science Teaching*, vol. 39, no. 5, pp. 410-422.

Scholtz, A 2007, 'An analysis of the impact of authentic assessment strategy on student performance in a technology-mediated constructivist classroom: A study revisited', *International Journal of Education and Development using Information and Communication Technology (IJEDICT)*, vol. 3, no. 4, pp. 42-53.

Schon, D 1983, *The reflective practitioner: how professionals think in action*, Area Publishing, Boston.

Scott, J 2000, 'Authentic assessment tools', in RL Custer (ed.), *Using authentic assessment in vocational education,* Center on Education and Training for Employment College of Education, Columbus, pp. 33-48.

Sellberg, C 2017, 'Simulators in bridge operations training and assessment: a systematic review and qualitative synthesis', *WMU Journal of Maritime Affairs*, vol. 16, no. 2, pp. 247-263.

Silvestri, L & Oescher, J 2006, 'Using rubrics to increase the reliability of assessment in health classes', *International Electronic Journal of Health Education*, vol. 9, pp. 25-30.

Spady, WG 1994, *Outcome-based education: Critical issues and answers*, American Association of School Administrators, Arlington, VA.

STCW 2011, *International convention on standards of training, certification and watchkeeping for seafarers*, International Maritime Organization, London.

Stellmack, MA, Konheim-Kalkstein, YL, Manor, JE, Masset, AR & Schmitz, JAP 2009, 'An assessment of reliability and validity of a rubric for grading APA-Style introductions', *Teaching of Psychology*, vol. 36, no. 2, pp. 102-107.

Stevens, P 2013, 'An examination of a teacher's use of authentic assessment in an urban middle school setting', PhD thesis, Ohio University.

Suresh, KP & Chandrashekara, S 2012, 'Sample size estimation and power analysis for clinical research studies', *Journal of Human Reproductive Sciences,* vol. 8, no. 3, pp. 186-192.

Tan, SH & Tan, SB 2011, 'The correct interpretation of confidence intervals', *Proceedings of Singapore Healthcare*, vol. 19, no. 3, pp. 276-278.

Tanner, DE 2001, 'Authentic assessment: A solution, or part of the problem?', *The High School Journal*, vol. 85, no. 1, pp. 24-29.

Tavakol, M & Dennick, R 2011, 'Making sense of Cronbach's alpha', *International Journal of Medical Education,* vol. 2, pp. 53-55.

Taylor, JM 2011, 'Interdisciplinary authentic assessment: Cognitive expectations and student performance', PhD thesis, Pepperdine University.

Tennant, M 1999, 'Is learning transferable?', in D Boud & J Garrick (eds.), *Understanding Learning at Work*, Routledge, London, pp. 165-180.

Teodorescu, T 2006 'Competence versus competency: What is the difference?', *Performance Improvement*, vol. 45, no. 10, pp. 27-30.

Theotokas, I & Progoulaki, M 2007, 'Cultural diversity, manning strategies and management practices in Greek shipping', *Maritime Policy & Management*, vol. 34, no. 4, pp. 383-403.

Thomas, J 2000, *A review of research on project-based learning,* viewed 16 September 2017, http://www.bie.org/research/study/review_of_project_based_learning_2000.

Todorov, KR & Brousseau, B 1998, *Authentic assessment of social studies*, Michigan Department of Education, Lansing, Michigan.

Trygstad, P 2010, 'Student engagement and student voices', PhD thesis, University of North Carolina.

Uchida, M 2004, 'Analysis of human error in marine engine management', *IAMU AGA 5: Advances in International Maritime Research*, *2004*, Australian Maritime College, Tasmania, Australia, pp. 85-93.

University of Tasmania (UTAS) 2011, *Guidelines for good assessment practice*, UTAS, Tasmania.

Varley, MA 2008, 'Teachers' and administrators' perceptions of authentic assessment at a career and technical education centre', PhD thesis, Fordham University.

Vendlinski, T, Underdahl, J, Simpson, E & Stevens, R 2002, 'Authentic assessment of student understanding in near-real time', NECC *2002: 23rd National Education Computing Conference Proceedings*, *June 17-19, 2002*, San Antonio, Texas, pp. 1-17.

Walczak, A 1999, 'STCW Convention and the challenges of the future', *Annual of Navigation*, vol. 1, pp. 21-27.

Walvoord, BE 2004, *Assessment clear and simple: A practical guide for institutions, departments and general education*, Jossey-Bass, San Francisco.

Wellington, P, Thomas, I, Powell, I & Clarke, B 2002, 'Authentic assessment applied to engineering and business undergraduate consulting teams', *International Journal of Engineering Education*, vol. 18, no.2, pp. 168-179.

Wiggins, G 1989, 'A true test: Toward more authentic and equitable assessment', *The Phi Delta Kappan*, vol. 70, no. 9, pp. 703-713.

Wiggins, G 1990, 'The case for authentic assessment', *Practical Assessment, Research and Evaluation*, vol. 2, no. 2, pp. 1-6.

Wiggins, G 1992, 'Creating tests worth taking', *Educational Leadership*, vol. 49, no. 8, pp. 26-34.

Wiggins, G 1993, 'Assessment: Authenticity, context and validity', *Phi Delta Kappa International*, vol. 75, no. 3, pp. 200-208,210-214.

Wiggins, G 1998, *Educative Assessment: Designing assessments to inform and improve student performance*, Jossey-Bass Publishers, San Francisco.

Williams, B, Brown, T & Onsman, A 2010, 'Exploratory factor analysis: A five-step guide for novices', *Journal of Emergency Primary Health Care*, vol. 8, no. 3, pp. 1-13.

Yong, AG & Pearce, S 2013, 'A beginner's guide to factor analysis: focusing on exploratory factor analysis', *Tutorials in Quantitative Methods for Psychology*, vol. 9, no. 2, pp. 79-94.

Zhang, L & Zheng, Y 2018, 'Feedback as an assessment for learning tool: How useful can it be?', *Assessment & Evaluation in Higher Education,* vol. 43, no. 7, pp. 1120-1132.