

# Authentic Versus Traditional Assessment: An Empirical Study Investigating the Difference in Seafarer Students' Academic Achievement

Samrat Ghosh<sup>1</sup>, Benjamin Brooks<sup>2</sup>, Dev Ranmuthugala<sup>1</sup>, and Marcus Bowles<sup>3</sup>

<sup>1</sup>(National Centre for Ports and Shipping, University of Tasmania Australian Maritime College, Launceston, Australia)

<sup>2</sup>(National Centre for Ports and Shipping, UTAS - Australian Maritime College, Launceston, Australia)

<sup>3</sup>(School of Engineering and Built Environment, Deakin University, Burwood, Australia)  
(E-mail: [sghosh@utas.edu.au](mailto:sghosh@utas.edu.au))

Past research showed that traditional assessment methods that required seafarer students to construct responses based on memorisation and analysing information presented in absence of real-world contexts (e.g. oral examinations and multiple-choice questions) disengaged the students from learning. Memorising information is a lower-order cognitive ability, failure in which led to errors and low academic achievement for students. Authentic assessment methods require students to construct responses through the critical analysis of information presented in real-world contexts. Hence, this research 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 (the 'control' group 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. Because some past researchers had defined and implemented traditional assessment methods as a single-occasion assessment, this project implemented the assessment in a summative format, as opposed to authentic assessments implemented during student preparation. Analysis of student scores revealed that the authentically assessed students were guided towards significantly higher academic achievement.

## KEY WORDS

1. Authentic assessment.
2. Traditional assessment.
3. Formative.
4. Summative.
5. Academic achievement.

Submitted: 14 April 2019. Accepted: 22 September 2019. First published online: 12 February 2020.

1. INTRODUCTION. Past research (AMC, 2011; Emad and Roth, 2007; Maringa, 2015) has demonstrated that seafarer students tended to disengage with traditional assessment methods (e.g. multiple-choice questions (MCQ), oral examinations, and written assignments) that were presented devoid of real-world contexts and focused only on their ability to recall and regurgitate the body of knowledge taught in the classroom. Disengaged students opt 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 method based on memorised answers in an oral examination. However, the ability to memorise is a lower-level cognition, and memory lapses may lead to unintentional skill and knowledge-based errors (Wiggins, 1989), leading to poor academic achievement. Although one may argue that traditional assessment methods like 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 regurgitation. Traditional assessment methods adopted in seafarer education are promoted by the Standards of Training, Certification and Watchkeeping (STCW) Code that was introduced by the International Maritime Organization (IMO) in 1978 (revised through major amendments in 1995 and 2010) to provide global minimum standards of competence assessment. The assessment methods may be effective in assessing knowledge-based components of a task, but 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).

Authentic assessment requires students to provide responses to a situation described and delivered in a real-world (or contextually similar) context (Wiggins, 1989). Authentic assessment tasks are found meaningful to students due to their strong figurative context and fidelity to the situations in the professional world (Wiggins, 1989). Meaningful tasks set in real-world contexts enhance student engagement with assessment if students relate the tasks to professional practices (Quartuch, 2011; Richards-Perry, 2011). Past research (Brawley, 2009; Gallagher et al., 1992; Leon and Elias, 1998; Thomas, 2000; Schneider et al., 2002) empirically proved higher student academic achievement for authentically assessed students when compared with their traditionally assessed counterparts. Similar evidence is essentially missing in the area of seafarer education. However, the focus of this paper was not only to validate previous findings in non-seafarer research but also to investigate how innovative assessment practices may be integrated into seafarer education to investigate its impact on the achievement of professional competence standards, as described by the STCW Code.

Hence, the objective of this research project was to investigate whether authentic assessment implemented in seafarer education significantly increased students' academic achievement (through the comparison of scores obtained) as compared with traditional assessments for a unit of competence listed in the STCW Code. Separate and independent seafarer student groups were identified as the control (traditional assessment) and the treatment group (authentic assessment). The assessments were implemented in the selected unit of 'Managerial and Leadership Skills' (listed as 'Use of leadership and managerial skill' in the STCW Code) 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 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, and, hence, maximised the number of participants.

The authentic assessment implemented differed from the traditional assessment on the basis of the inclusion of a real-world context that attempted to closely replicate the real-life 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 (Abeywickrama, 2012; Bailey, 1998, p. 205; Dikli, 2003, p. 16; Law and Eckes, 1995), traditional assessment methods have been conventionally described as not only inauthentic but also as a 'one-shot' and single-occasion tests implemented at the end of the 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). 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). In comparison to the summative traditional assessments, one of the key characteristics of authentic assessment, as defined by its major authors (Gulikers, 2006; Wiggins, 1989), required students to be informed of 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 and Eckes, 1995) before receiving the final judgement on their competence.

Hence, the authentic assessment in this research project was implemented as a formative assessment and 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; this research also investigated the difference in 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 their effect on academic achievement. The findings of this research project revealed recommendations for seafarer education providers towards the implementation of authentic assessment and improvement of students' academic achievement.

## 2. RESEARCH METHODOLOGY.

2.1. *Research design.* The difference in students' academic achievement 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 on board a merchant ship. The students who enrolled in the unit in Semester 1 were classified as the 'control group' that underwent a traditional assessment. The traditional assessment comprised two case study scenarios 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 on board ships.

In comparison, another cohort of students enrolled in the same unit in Semester 2 were assessed authentically through the same case studies 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 life raft during a fire was demonstrated at the AMC training pool. The pool was equipped with facilities to launch a real life raft in simulated waves, strong winds, darkness, rain, and smoke. The simulation also included sounding of the emergency alarms and staff playing the role of panicking seafarers jumping into the pool to replicate a possible emergency. Although the focus of the authentic assessment was also to assess the students' ability to meet the learning outcomes through written answers, the students were able to engage themselves in the sensory experience of the demonstrated case studies. In comparison to the authentic assessment, students assessed traditionally had to rely only on their imagination and experience to visualise the described scenarios.

Although it may be argued that the descriptive case studies in themselves (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 (three weeks apart). The second authentic task was implemented once the students received individual feedback on their performance in the first authentic task. 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. A minimal-risk ethics application approval was obtained for this research project. [Table 1](#) summarises the research design.

**2.2. 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 of the difference in students' academic achievement between the two types of assessments

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
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	1 h	30 min for each case study (Total: 1 h)

implemented, standard deviation (SD) 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 ( $P < 0.05$ ) indicated whether the variation in scores of students' academic achievement was statistically significant for reporting.

2.3. *Sampling considerations.* 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). 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 those students to complete the administrative paper work. A key consideration during sampling was to ensure that the control and treatment groups comprised 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 predetermined characteristics, such as qualifications, academic ability or work experience, that may have impacted the outcomes of this research.

#### 2.4. *Validity and reliability of assessment.*

2.4.1. *Before implementing assessment.* Content and construct validity were achieved by using a jury of experts before the assessment was implemented. The subject experts comprised 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.

2.4.2. *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.

### 3. RESEARCH QUESTIONS, HYPOTHESIS AND INDEPENDENT VARIABLES.

The focus of this research project led to the following research questions (RQ):

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

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

1a)  $Score AA > Score TA$

As stated in Section 2.1, 'Research design', 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 that defined the standards and criteria of performance achieved by the students. In addition to the rubric, feedback comments were included in the students' answer sheets for their perusal. Finally, a generalised feedback was 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)  $H_{1b}$ : *Score of the first authentic assessment task ( $AA_1$ ) > Score of the first traditional assessment task ( $TA_1$ )*

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 individual feedback. Thus, apart from the 'authentic' design, an additional variable (an opportunity to improve achievement in authentic assessment) that may have influenced student achievement in this research was introduced due to the nature of assessment implementation. 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)  $H_{2a}$ : *Score of the second authentic assessment task ( $AA_2$ ) > Score of the second traditional assessment task ( $TA_2$ )*

Table 2. Research questions and the resulting hypothesis.

Research Question	Hypothesis
<b>RQ1:</b> Is there a significant improvement in seafarer students' academic achievement in authentic assessment when its scores are compared with traditional assessment scores?	<b>H<sub>1a</sub>:</b> Score AA > Score TA <b>H<sub>1b</sub>:</b> Score AA <sub>1</sub> > Score TA <sub>1</sub>
<b>RQ2:</b> Is there a significant improvement in seafarer students' academic achievement in formative authentic assessment when its scores are compared with summative traditional assessment scores?	<b>H<sub>2a</sub>:</b> Score AA <sub>2</sub> > Score AA <sub>1</sub> <b>H<sub>2b</sub>:</b> Score AA <sub>2</sub> > Score TA <sub>2</sub> <b>H<sub>2c</sub>:</b> Score TA <sub>2</sub> ~ Score TA <sub>1</sub>

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 who 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:

$$2b) H_{2b}: \text{Score } AA_2 > \text{Score } AA_1$$

Since the summative nature of the traditional assessment did not allow students to receive individual feedback 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:

$$2c) H_{2c}: \text{Score } TA_2 \sim \text{Score } TA_1$$

The research questions and the resulting hypotheses is summarised in [Table 2](#).

3.1. *Independent variables.* This research identified the independent variables that could influence the academic achievement. Since the assessments required 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 assessments 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

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.
RQ2	$H_{1b}$ : Score AA <sub>1</sub> > Score TA <sub>1</sub>	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 <sub>2</sub> > Score TA <sub>2</sub>	Student achievement was significantly higher in authentic assessment for the composite group and groups isolated on independent variables.
RQ2	$H_{2b}$ : Score AA <sub>2</sub> > Score AA <sub>1</sub>	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 <sub>2</sub> ~ Score TA <sub>1</sub>	No significant difference in seafarer student achievement found for the composite group and groups isolated on independent variables.

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 undergraduate or postgraduate 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. RESULTS. The results were analysed against the RQs and corresponding hypothesis described previously. Findings are summarised in [Table 3](#).

The results are presented below.

- **RQ1:** AA significantly improved by 17.3 per cent compared with TA; and AA<sub>1</sub> was 11.4 per cent higher than TA<sub>1</sub>. The hypotheses ( $H_{1a}$  and  $H_{1b}$ ) designed for RQ1 thus held true. In both hypotheses, the SD 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:** Analysis of the composite scores showed the following:
  - AA<sub>2</sub> significantly improved by 23.2 per cent when compared with TA<sub>2</sub>
  - AA<sub>2</sub> significantly improved by 12 per cent when compared with AA<sub>1</sub>
  - No significant difference was found between TA<sub>1</sub> and TA<sub>2</sub>



The hypotheses ( $H_{2a}$ ,  $H_{2b}$  and  $H_{2c}$ ) designed for RQ2 thus held true. In hypothesis  $H_{2a}$  and  $H_{2b}$ , the SD, effect size and  $t$ -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 SD 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.

4.1. *The effect of independent variables on students' academic achievement.*  $H_{1a}$ ,  $H_{2a}$  and  $H_{2b}$  held true for all the independent variables. The SD, effect size and  $t$ -test values showed that the difference and variation in scores was significant for reporting for all the independent variables.  $H_{1b}$  held true for all the independent variables but with a single exception. 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 SD, effect size and  $t$ -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 SD 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.

$H_{2c}$  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 per cent, 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 SD values indicated that the scores of both the traditional tasks were similarly scattered, and the recommended effect size and the  $t$ -test values showed that the difference and variation in scores was insignificant for reporting.

## 5. DISCUSSION.

5.1. *Higher academic achievement in authentic assessment.* The results of this research confirmed 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 imaginings of a situation but on the assimilation, integration and analysis of information provided in a real-world context. This finding corroborated the findings of non-seafarer research (Brawley, 2009; Gallagher et al., 1992; Leon and Elias 1998; Schneider et al., 2002; Thomas, 2000) 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. For example, Brawley (2009) compared academic achievement of early childhood students, Schneider et al. (2002) 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 level school students.

This research also distinguished itself from past research by using two separate student groups as the 'control' and 'treatment' groups. 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 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 this research.

Although Schneider et al. (2002), Thomas (2000), Leon and Elias (1998) and Gallagher et al. (1992) used two separate randomly assigned groups for comparison between authentic and traditional assessment, 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 'performance-based projects'; Gallagher et al. (1992) used 'open-ended questions' versus 'authentic task'; and Schneider et al. (2002) and Thomas (2000) used two separate groups with a different learning experience.

5.2. *Higher academic achievement in formative assessment.* The formative assessment employed in this research project provided students with an opportunity to receive individual feedback on their performance in AA<sub>1</sub> before attempting AA<sub>2</sub>. According to Zhang and Zheng (2018), providing feedback on a students' current ability to perform an assessment task and making 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<sub>2</sub> as compared with AA<sub>1</sub> indicated that, using the feedback obtained, seafarer students recognised the gaps in their knowledge, reevaluated 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 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 is feedback on the students' current level of skills and multiple opportunities to practice those skills. This was also empirically reconfirmed in this project and is evident in the scores of students in the comparison of AA<sub>1</sub> to TA<sub>1</sub> and AA<sub>2</sub> to TA<sub>2</sub>, when isolated for the independent variable of 'more than three years of work experience'. The scores of AA<sub>1</sub> were similar to those of TA<sub>1</sub> when isolated for the specified variable which indicated that 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<sub>2</sub> were significantly higher than those of TA<sub>2</sub> 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 feedback and an opportunity to improve on their performance and resulting scores.

5.3. *Impact of independent variables on academic achievement in authentic assessment.*

5.3.1. *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 student achievement was significantly higher for students with more than three years of work experience as compared with students with fewer than three years of work experience. This implied that students with less 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 score significantly 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.

5.3.2. *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 groups 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 before educators is facing the challenge of teaching and assessing students with differing abilities in communicating using the English language. 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 higher standards.

However, based on the current scenario of a classroom with diverse population of seafarer students with differing abilities to communicate in the English language, educators must investigate ways to design authentic assessments that require students to perform tasks that require more a 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.

5.3.3. *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 a university education did not score significantly higher (a comparison of means indicated that the difference of marks was

less than 2 per cent) than the students with only high school qualifications. This finding indicated that authentic assessment enacted in real-world settings equalised student ability in analysing contexts 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 assist in the retrieval of information and prevent lapses in memory (Atkinson and 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.

By comparison, analysis of the scores within the control group (overall and when isolated on independent variables) revealed that students with university-level qualification scored significantly higher than did students with only high school qualifications. In absence of a real-world contexts, students in the 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 higher than did their less-educated (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 undergraduate and postgraduate qualifications that 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 gaps between students with differing educational backgrounds.

5.4. *Disadvantages of implementing authentic assessment.* Although the implementation of authentic assessment in this project provided a significant advantage through higher academic achievement for 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 authentic assessments are time consuming and cost intensive. Hence, a comparative cost and time analysis for the resources used in developing the traditional and authentic assessment was conducted for this research. Table 4 details the differences in the resources used and the costs incurred.

In Table 4, the time and costs (shown in Australian dollars) analysis conducted for this project showed that 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 et al. (2017) argued that changes in the assessment regime become justified if the benefits outweigh the costs. In this project, the benefit obtained through higher academic achievement for authentically assessed students justifies this assessment regime implementation in seafarer education. Also, better educated and knowledgeable seafaring students that have achieved improved professional competence will reduce financial costs that shipping companies incur due to human error (Rothblum,

Table 4. Comparison of cost estimation in assessment implementation (AMC Business Manager, 2018).

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

2000). 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 that their employees are competent to work on board ships (IMO, 2002). Ships can be detained, and registers cancelled, if serious deficiencies are found in an operator's ability to perform workplace tasks safely. Competent seafarers recruited on ships will have the potential of enhancing ships' turnaround time, meeting the efficiency demands of ship owners and safety performance of ship operations (Yuen et al., 2018), which could potentially translate to sizeable cost savings and service improvements for a shipping company.

6. CONCLUSION. This research project was a rigorous experiment designed 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, any independent variables which could possibly affect student performance were 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 students assessed authentically had significantly higher scores than did those students assessed traditionally, resulting in higher academic achievement. This finding indicated that student academic achievement will be improved if students focus on the assimilation, critical analysis and integration of information presented through a real-world context instead of by memorising information and rote learning.

The findings of this research indicated that student academic achievement will be improved if they are provided with feedback that may be used in recognising 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. Hence, in the context of seafarer education, a shift is required from summative oral assessments which 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 student performance and is often too late to make any changes to the extant learning strategies. However, educators should provide timely and efficient feedback to students, and receive counter feedback to reflect on areas where they may improve 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 recognising their strength and weaknesses; and in establishing realistic learning goals. This develops their metacognitive ability to reflect on their current learning practices and improve on them. Reflection on practices is a critical part of professional performance required to avoid repeated 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, for educators working together, policies of education institutes must support the organisation 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 more work experience, proficiency in the English language and advanced educational qualifications. To address the needs of the learners with different backgrounds and to achieve equity in academic achievement, this project recommends that educators 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.

Future research should investigate factors of assessment (task, context, 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 with their scores in the assessment task. The factors correlating significantly will be included in designing assessment tasks towards improving the academic achievement.

## REFERENCES

- Abeywickrama, P. (2012). Rethinking traditional assessment concepts in classroom-based assessment. *The CATESOL Journal*, *23*(1), 205–214.
- AMC (Australian Maritime College). (2011). *Report on the Review of the National Centre for Ports & Shipping: Deck Officer Courses*. Tasmania: AMC.
- AMC (Australian Maritime College) Business Unit Manager. (2018). *Costs and Pricing*. Confidential and controlled document copy. Tasmania: AMC.
- Atkinson, R., and Shiffrin, R. (1968). Human memory: A proposed system and its control processes. In: Spence, K., and Spence, J. (eds.), *The Psychology of Learning and Motivation* (Vol. 2). New York: Academic Press, 89–195.
- Bailey, K. M. (1998). *Learning about Language Assessment: Dilemmas, Decisions, and Directions*. New York: Heinle & Heinle.
- Boud, D., and Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, *31*(4), 399–413. doi: 10.1080/02602930600679050.
- Brawley, N. (2009). Authentic assessment vs traditional assessment: A comparative study. Honours thesis, Coastal Carolina University, Conway, SC.
- Clark, R. A. (2014). Correlation study: *The effect of student-teacher rapport on high school student performance rate*. Doctoral dissertation, Liberty University, Lynchburg, VA.
- 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. London: University of Exeter.
- Dikli, S. (2003). Assessment at a distance: Traditional vs. alternative assessments. *The Turkish Online Journal of Educational Technology*, *2*(3), 13–19.
- Emad, G., and Roth, W. M. (2007). Evaluating the competencies of seafarers: Challenges in current practice. In: Pelton, T., Reis, G., and Moore, K. (eds.). *Proceedings of the University of Victoria Faculty of Education Research Conference - Connections '07*. Victoria, BC: University of Victoria, 71–76.
- Gallagher, S. A., Stepien, W. J., and Rosenthal, H. (1992). The effects of problem-based learning on problem solving. *Gifted Child Quarterly*, *36*(4), 195–200. doi: 10.1177/001698629203600405.
- Gulikers, J. T. M. (2006). *Authenticity is in the eye of the beholder: Beliefs and perceptions of authentic assessment and the influence on student learning*. Doctoral dissertation, Open Universiteit Nederland, Limburg, Netherlands.
- Hattie, J. A. C. (2009). *Visible learning: A Synthesis of over 800 Meta Analyses Relating to Achievement*. London: Routledge.
- IMO (International Maritime Organization). (2002). *International Safety Management (ISM) Code and Revised Guidelines on Implementation of the ISM Code*. London: IMO.
- Joughin, G., Dawson, P., and Boud, D. (2017). Improving assessment tasks through addressing our unconscious limits to change. *Assessment & Evaluation in Higher Education*, *42*(8), 1221–1232. doi:10.1080/02602938.2016.1257689.
- Law, B., and Eckes, M. (1995). *Assessment and ESL: On the Yellow Big Road to the Withered of Oz*. Manitoba: Peguis.
- Leon, S., and Elias, M. (1998). A comparison of portfolio, performance, and traditional assessment in the middle school. *Research in Middle Level Education Quarterly*, *21*(2), 21–37. doi: 10.1080/10848959.1998.11670118.
- Maltby, A., and Mackie, S. (2009). Virtual learning environments – help or hindrance for the 'disengaged' student? *ALT-J*, *17*(1), 49–62. doi:10.1080/09687760802657577.
- Maringa, M. T. H. (2015). *Assessment of quality of training and education of seafarers in South Africa and Ghana*. Master's thesis, World Maritime University, Malmö, Sweden.

- Mueller, J. (2006). *Authentic assessment toolbox*. Retrieved from <http://jonathan.mueller.faculty.noctrl.edu/toolbox/whatisit.htm#looklike>
- Neely, P., and Tucker, J. (2012). Using business simulations as authentic assessment tools. *American Journal of Business Education*, **5**(4), 449–456.
- Prasad, R. (2011). *The role of collaborative learning in promoting safety in shipboard machinery spaces*. Doctoral dissertation, World Maritime University, Malmo, Sweden.
- Quartuch, M. J. (2011). *Is authentic enough? Authentic assessment and civic engagement*. Master's thesis, Moravian College, Bethlehem, PA.
- Richards-Perry, G. D. (2011). *Student perceptions of engagement in schools: A Deweyan analysis of authenticity in high school classrooms*. Doctoral dissertation, Georgia State University, Atlanta, GA.
- Rothblum, A. M. (2000). Human error and marine safety. Proceedings of the Maritime Human Factors Conference. Linthicum: Maryland, USA, 1–10.
- Schneider, R. M., Krajcik, J., Marx, R., and Soloway, E. (2002). Performance of students in project-based science classrooms on a national measure of science achievement. *Journal of Research in Science Teaching*, **39**(5), 410–422. doi:10.1002/tea.10029
- Thomas, J. (2000). *A review of research on project-based learning*. Retrieved from [http://www.bie.org/research/study/review\\_of\\_project\\_based\\_learning\\_2000](http://www.bie.org/research/study/review_of_project_based_learning_2000)
- Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *The Phi Delta Kappan*, **70**(9), 703–713. doi:10.1177/003172171109200721.
- Yuen, K. F., Loh, H. S., Zhou, Q., and Wong, Y. D. (2018). Determinants of job satisfaction and performance of seafarers. *Transportation Research Part A: Policy and Practice*, **110**, 1–12.
- Zhang, L. and Zheng, Y. (2018). Feedback as an assessment for learning tool: How useful can it be? *Assessment & Evaluation in Higher Education*, **43**(7), 1120–1132. doi:10.1080/02602938.2018.1434481.