Journal of King Saud University – Engineering Sciences (2012) **24**, 95–105

King Saud University

Journal of King Saud University – Engineering Sciences

www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

The reliability and validity of Basic Offshore Safety and Emergency Training knowledge test

Mohamad Fahmi Hussin [a,*], Bin Wang [a], Ramani Hipnie [b]

[a] *Safety Engineering Department, Aberdeen University, AB24 3UE Aberdeen, Scotland, United Kingdom*
[b] *Construction and Industrial Safety Training Center (CONSIST), 68000 Ampang, Selangor, Malaysia*

Received 17 October 2010; accepted 10 May 2011
Available online 24 August 2011

KEYWORDS

Knowledge retention;
B.O.S.E.T;
Risk management;
Safety training

Abstract This paper presents the development of a Basic Offshore Safety and Emergency Training (B.O.S.E.T) knowledge test. The knowledge test was developed to measure B.O.S.E.T knowledge retention among offshore professionals. This research requires the knowledge test to be administered every two months, in a period of six months. The objective of this paper is to present key points that validate the research tool in terms of readability and validity. Three readability tests (Flesch-Kincaid Grade Level Test, Flesch Reading Ease Test and Gunning's Fog Index Test) were used to identify the suitability of test language. In addition, the knowledge test was subjected to face validity and content validity. Seventy-nine B.O.S.E.T trainees took part in this research. The test results suggest that the contents and the language used on the knowledge test is suitable for target sample; hence the test can be used to identify knowledge retention among offshore professionals.

1. Introduction

There are many risks and hazards of working in the offshore industry. These offshore risks and hazards include fire explosions, helicopter ditching, boat capsize and many more. When such an accident does occur offshore, the situation becomes worse knowing that any form of rescue or help may take more time and effort to arrive. Thus, managing risk and hazards is an important aspect for safety in the offshore industry.

One of the methods of managing offshore risk is by ensuring offshore workers are equipped with offshore safety knowledge and skills. According to Hubbard (2009), offshore risk management refers to identifying and assessing offshore risks and identifying solutions to control or minimize the probability or the impact of such risk. In addition, Hubbard stated that while technological solutions have vastly improved offshore safety in many ways, more effort is needed to improve human capacity solutions. Human capacity solution plays an important role in managing offshore risks and hazards (Hubbard, 2009). An example of human capacity solutions is training; in this case – Basic Offshore Safety and Emergency Training (B.O.S.E.T).

* Corresponding author.
E-mail addresses: Mohamad.fahmi.hussin@abdn.ac.uk (M.F. Hussin), Bin.wang@abdn.ac.uk (B. Wang), Ramanie@consist.com.my (R. Hipnie).

Peer review under responsibility of King Saud University.
doi:10.1016/j.jksues.2011.05.002

ELSEVIER Production and hosting by Elsevier

The general offshore safety training is called Basic Offshore Safety and Emergency Training (B.O.S.E.T). The B.O.S.E.T course is made of three topics – Offshore Safety Induction and Fire Fighting, Sea Survival and Helicopter Underwater Egress Training (H.U.E.T). The B.O.S.E.T course is implemented around the around and it is a pre-requisite upon enrolling into a B.O.S.E.T course that a trainee is required to produce an offshore medical certificate; which can be obtained from medical doctors. Doctors will award a medical certificate based on results from physical tests such as eye vision test, hearing test and lung capacity check. The medical certificate validates a person for a period of two (2) years. After the two (2) year period, the medical certificate becomes invalid and trainees are required to re-do the medical certificate again.

The objectives of the B.O.S.E.T course include to increase offshore workers survivability by understanding techniques for signaling, rescue and survival in the open ocean, familiar with survival situations in cold water as well as to obtain experience in the care, donning and use of immersion suits and survival swimming. In addition, the B.O.S.E.T course prepares offshore workers with knowledge of personal floatation device used in an emergency, techniques of vessel and platform abandonment and familiarity to marine life support equipment and its functions. Finally, the B.O.S.E.T course works to control offshore workers anxiety over real emergencies by participating in realistic underwater simulations, and through the use of Personal Protective Equipment (PPE) (Mior and Ramanie, 2009).

The objective of this research is to identify B.O.S.E.T knowledge retention among offshore professionals. However, this paper will only discuss the key points that validate the questionnaire used to identify the B.O.S.E.T knowledge retention.

2. B.O.S.E.T knowledge test

The B.O.S.E.T knowledge test was developed in four different stages. These stages were:

(a) Define B.O.S.E.T knowledge test scope.
(b) Define target population.
(c) Consultations with experts.
(d) Pilot tests.

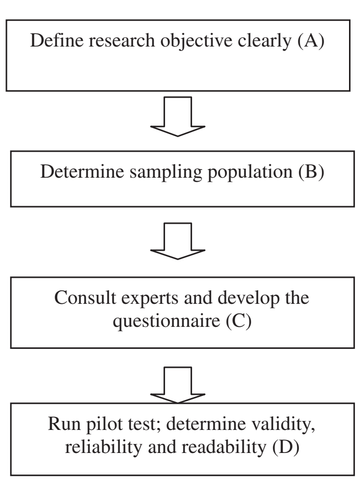

Figure 1 Knowledge test design outline. Adapted from Trochim (2005), Kumar (2005), Bradburn et al. (2004), Schwarz and Oyserman(2001).

Fig. 1 illustrates the four stages in developing the B.O.S.E.T knowledge test.

2.1. Scope of Knowledge test

All of the questions in the B.O.S.E.T knowledge test revolve around the three main B.O.S.E.T course topics [Safety Induction and Fire Fighting, Sea Survival and Helicopter Underwater Egress Training (H.U.E.T)]. Fourteen questions were constructed for the test, of which three questions were from Fire Fighting and Safety Induction, five questions were refer to Sea Survival and six questions on H.U.E.T. The guideline for the development of the B.O.S.E.T knowledge test and the objective and reasoning behind the knowledge test questions are made available in Appendix A.

2.2. Target population

The second stage of the tool design involves identifying a sampling population or a group of people suitable for this research. This study defines the research participants as Class A offshore professionals. Here a professional is referred as someone who does a job that requires special training, education or skills, for example, electrical engineers. The research focuses on three groups of professionals:

(a) Engineers.
(b) Supervisors.
(c) Managers.

These three groups of professionals were chosen because their associated responsibilities to ensure safety, health and welfare of the working force offshore.

2.3. Consultations with experts

The next stage of the questionnaire design was the consultation with the B.O.S.E.T experts in regards to the knowledge test. An expert is someone who has extensive knowledge of skills in a particular area. Since this research revolves around the B.O.S.E.T training course, the research consultation was carried out with the course providers. Difficulties had been encountered with the training centers in Aberdeen area due to data sensitivity. As the result, consultation and test were carried in Malaysia offshore industry which is more open and accommodating for research in this nature.

The first step taken in identifying offshore training centers in Malaysia was contacting Petroliam Nasional Berhad (PETRONAS), Malaysias' national petroleum company. It was made known that PETRONAS sends their workers to the Construction and Industrial Safety Training Centre (CONSIST). Followed up with a discussion, the chairman of CONSIST training centre, Mr. Mior Ar Zawari, kindly agreed to allow the research to be facilitated at CONSIST as well as to contribute their expertise in the research. The manager of CONSIST is Mr. Ramani Hipnie who has over 20 years experience working in the offshore industry in Australia and Canada with Survival Systems. This research refers to Mr. Ramani and his team of B.O.S.E.T trainers as experts for consultations and guidance.

2.4. Pilot tests

The research ran two (2) pilot tests on the knowledge test to detect any possible ambiguities present in the test as well as to evaluate the time taken by candidates to answer all the questions (Trochim, 2005; Kumar, 2005). The first pilot test was conducted in early September 2009 with 20 B.O.S.E.T trainees from CONSIST. Participants were asked for their interpretations of the questions in the designed test. Suggestions and feedbacks from participants were considered in improving the knowledge test. For instance, in some cases, there were difficulties in choosing the appropriate wording for the questions, thus, reference to the B.O.S.E.T manual was added to the questions as improvement. In the second pilot test, an evaluation of the average time required to answer the knowledge test was conducted. Due to the limited ability of offshore professionals to commit time for the research, an answering timescale had to be set. An analysis of the pilot test indicates that participants can accommodating to commit between 15 and 20 min to answer the knowledge test (Wells and Wallock, 2003). Based on these pilot studies, the knowledge test was designed and Fig. 1 shows the flow chart in the overall test design procedure.

2.5. Characteristics of the B.O.S.E.T knowledge test

The knowledge test was designed to have three important characteristics:

(a) The knowledge test is ‘readable’ for the targeted population sample – the knowledge test must be able to demonstrate that participants understood the questions in the test (Paul, 2003; Klare, 2000).
(b) The knowledge test is able to differentiate trained and untrained participants. The knowledge test must be able to demonstrate the ability to distinguish between these two groups.
(c) The knowledge test is ‘highly discriminated/ finely tuned’ – a ‘highly discriminated’ or ‘finely tuned’ knowledge test suggests that the test had been refined and minimizes effects of ambiguity for the questions in the tests (Milne, 1999).

2.6. Ethical consideration

The researcher acknowledges the importance of ethical considerations when there is human interaction involved. The research strives to protect participants’ rights before, during and after the conduct of the study. The research used the principle of voluntary participation, whereby the participants were not coerced into participating in the research (Trochim, 2005). A consent form was developed and administered to all participants before the start of the B.O.S.E.T course. It informs the participants with all relevant details in regards to the research and validates ethical considerations for both knowledge test and skills test.

2.7. Administering the knowledge test

To identify the B.O.S.E.T knowledge retention, the knowledge test was administered five times on the following occasions:

(a) Before the B.O.S.E.T course (pre-course).
(b) After the B.O.S.E.T course (post course or 0 month).
(c) Two (2) months after the B.O.S.E.T course (2 months).
(d) Four months after the B.O.S.E.T course (4 months).
(e) Six months after the B.O.S.E.T course (6 months).

The concept of repeatedly administering the knowledge test to the same research participants was based on previous knowledge and skills retentions studies. Berden et al. (1993) studied resuscitation skills retention among professional nurses through test repetition of three months, six (6) months and 12 months. Cullen (1992) studied knowledge retention among First Aiders in periods of six months, 12 months, 18 months, 24 months and 30 months. Other researchers such as Collins (2000), Curry and Gass, 1983, Fossel et al. (1983) and Deliere and Schneider (1980) have also identified retention model through repetitive testing method.

There were little or no interactions between the researcher and the research participants when the knowledge test was personally administered. By personally administering the knowledge test, this reduces the chance of ‘experimenter effect’ from occurring (Trochim and Donnelly, 2007). Experimenter effect occurs when a researcher or an experimenter unconsciously cues a number of subtle and non verbal or verbal signals to research participants which may influence or affect the research outcome. An example of non verbal cues includes body gestures such as the hand gestures; while an example of verbal signals include the tone of the voice or intonations. According to Rosenthal (1998), researches have successfully demonstrated that research experimenters may unintentionally communicate research expectation or objectivity through subtle cues, and this can significantly affect the research outcome. Therefore, the research had personally administered the knowledge test to the research participants as means of reducing the chance of experimenter effect.

3. Tool tests

The knowledge test was subjected to readability tests and validity tests. The readability test works to identify the suitability of the language used for the target research samples. In other words, the readability test indicates whether the wordings in the knowledge test could be understood by offshore professionals. If the wordings in the knowledge test proved difficult for participants to comprehend, this may suggest elements of ambiguities in the knowledge test. Thus, successful readability tests results imply that there are minimum comprehension ambiguities in the knowledge test. If the knowledge test demonstrates good readability test results, the research can be confident that research participants understood the questions and answers whole heartedly and with minimum effects of ambiguities.

In addition to that, the knowledge test had also undergone two (2) validity tests; face validity and content validity. Validity tests were used to ensure that the knowledge tests measures what it is supposed to measure. The knowledge test was developed to measure B.O.S.E.T knowledge among offshore professionals. It is important to conduct validity tests on the knowledge tests as the tests indicate the tool success in measuring desired data.

3.1. Readability test

When a research revolves around forms of writing, one of the major challenges is to measure comprehension difficulty. The tools used in this research must be able to demonstrate clear and concise language so that participants and easily understand the questions/scenarios. In this research, readability can be defined as the ease of which the research tool (knowledge test) can be read. There are mathematical formulas in Flesch Kincaid Grade Level test, Flesch Ease Reading test and Gunning Fog Index which are designed to assess the readability of a written document or written research tool (Paul, 2003; Klare, 2000).

The readability test has two main advantages for use. For document writers, the readability tests helps to assess the document complexity in a short period of time. If the document written is too complex, the readability test can 'screen' and minimize dense part of the document. In addition, the readability test can be used to identify the improvements made in a written document. The test provides a 'quantifiable' measure of improvements made in a written document (Fry, 2002; Hargis, 2000).

However, there are several issues that the readability tests are unable to deal with. The readability tests do not work to identify the attractiveness of written document nor does it identify the appropriateness of vocabulary used in the document. In other words, a written work can be poor but readable at the same time. The readability test does not work to identify differences in gender, class, religion or cultural, hence this may lead to illogical order of a written document. The use of readability test should only be for evaluating the ease of reading a document. This research uses tool readability to ensure that the tool is easily read and understood by the participants (Redish, 2000; Zakaluk and Samuels, 1988; Darville and Hiebert, 1985).

3.1.1. A Flesch-Kincaid Grade Level test

One of the ways to measure readability is by using Flesch-Kincaid readability test. The Flesch-Kincaid readability test is divided into Flesch Reading Easiness and Flesch Kincaid Grade Level. Both tests use the same core measuring method but different weighing factors. The formula for the Flesch Kincaid Grade Level test is given by:

$$\text{Grade Level Score} = 0.39\ (\text{total words}/\text{total sentences}) + 11.8(\text{total syllables}/\text{total words}) - 15.59$$

The test works to score readability into a United States grade level, thus making it easier to judge the readability of a text. In other words, the formula is able to estimate the number of US education years required to understand a text (DuBay, 2006; Wegner and Girasek, 2003). For example, if the formula generates a score of 7.3, this mean that the text is expected to be understandable by an average 7th grade (12–13 years old). Below is the USA education grade level table with respect to age (Table 1).

Table 1 US education grade level and age range (DuBay, 2006; Wegner and Girasek, 2003)

Grade	Age range (years old)
1st	6–7
2nd	7–8
3rd	8–9
4th	9–10
5th	10–11
6th	11–12
7th	12–13
8th	13–14
9th	14–15
10th	15–16
11th	16–17
12th	17–18

3.1.2. The Flesch Reading Ease test

The Flesch Reading Ease test is another readability test that was used in the research. It is slightly different as compared to Flesch Kincaid Grade Level test in the sense that the scores are rated 0–100. Unlike Flesch Kincaid Grade Level test, the higher the scores obtained from the Flesch Reading Ease test, the easier it is for the text to be understood (DuBay, 2006; Wegner and Girasek, 2003). The scores can be interpreted in Table 2.

Table 2 Flesch reading ease score grade (DuBay, 2006; Wegner and Girasek, 2003).

Score	Level	Equivalent US education grade
90–100	Very easy	5th
80–90	Easy	6th
70–80	Fairly easy	7th
60–70	Standard	8th or 9th
50–60	Fairly difficult	10th to 12th
30–50	difficult	13th to 16th
0–30	Very difficult	College grade

The formula for the Flesch Kincaid Readability is:

$$\text{Reading Ease Score} = 206.835 - 1.015\ (\text{total words}/\text{total sentences}) - 84.6\ (\text{total syllables}/\text{total words})$$

If the reading ease score of a written work yields a value 49, this mean that the work can be read by people who are above 13th grade. Another example is that a written work of reading ease score value 11 would be suitable for those who had gone to college. As a reference note, Reader's Digest magazine has a reliability index near to 65 while the Time magazine has a 52 score value.

3.1.3. The Gunning Fog Index test

In addition to the Flesch-Kincaid Grade Level Reliability test, the research tool was also subjected to the Gunning Fog Index (FOG) Readability test. The test was developed to eliminate unnecessary complexity in many literatures (full of 'fog'). The Gunning Fog Index requires the text to be at least 100 words long. The formula for the FOG index is given by:

$$\text{Gunning Fog Index} = 0.4\ [(\text{number of words}/\text{number of sentences}) + \text{Percentage of Hard words}]$$

Hard words are defined as words with three or more syllables such as comparable, fortunate and necessary. Generally, the FOG index should measure tool readability between scores of 7 and 8. If the FOG index is above 12, it is considered difficult. For example, Newsweek and Wall street Journal has an average FOG index of 11 while the bible has a FOG index of 6. The FOG index offers to simplify written work for easier comprehension. However, the FOG index also encourages low writing styles which may indirectly result in dull and uninteresting text written (DuBay, 2006; Harvey and Fleming, 2003).

3.2. Tool validity

Validity is an important tool element that measures a research's success at what the researcher sets out to do. In other word, the tool validity refers to the degree in the research tool that is truly measuring what it is intended to measure. Without tool validity, doubts may arise from the accuracy of data presented as well as the conclusions derived (Ogince et al., 2007; Li, 2003). The knowledge and skills test were subjected to internal validity check (Allen and Yen, 2002; Lacity and Jansen, 1994). Internal validity can be defined as the rigorous process taken into consideration while the research was conducted. The knowledge and skill tests were subjected to two (2) types of validity tests: the face validity and the content validity. This research recognizes the use of tool validity to measure the success of the research in obtaining the evidence and data.

3.2.1. Face validity

Face validity can be defined as a test that 'appears' as if it is going to measure what it is supposed to measure (Shuttleworth, 2009). Another way of defining face validity is a test that appears to be valid or accepted by the researcher or field experts (Schmitt and Landy, 1993). For this research, face validity was used by discussing the knowledge and skill tests with B.O.S.E.T experts, with the intention of obtaining views and suggestions for better tool modification. Six B.O.S.E.T experts went through and discussed the test items and were satisfied with both knowledge test and skill tests. However, the disadvantage of face validity is that there is a possibility for the experts to take their knowledge for granted as well as to overestimate how much people know or understand. This may result in the test being extremely difficult for participants to answer (Trochim, 2005; Kumar, 2005). However, it can be argued that all the questions and scenarios in the research tool revolve around the basic safety knowledge and skills that are expected for offshore professionals to retain after they have successfully completed the B.O.S.E.T training course.

3.2.2. Content validity

For this research, content validity is defined as validation of research tool based from credible resources such as the B.O.S.E.T manual. The research used content validity to validate both the knowledge test and the skills test. This means that the items in the knowledge test and skills test represent only a small slice of a larger domain of knowledge and skills. There were two (2) issues considered important in the content validity; the source of content and the professional acceptance (McKnight, 1999).

3.2.2.1. Source of content. The knowledge test was only administered to those who had enrolled in B.O.S.E.T course. CONSIST provides the B.O.S.E.T manual to all its trainees. The B.O.S.E.T contains all relevant safety information that the participants is expected to acquire. The contents in the knowledge test and the skill tests were referenced to the B.O.S.E.T manual (Li, 2003; Kane, 2001).

3.2.2.2. Professional acceptance. To all professionals taking part in the research, content validity can practically be defined by the content of the manual they used during the B.O.S.E.T course. This definition is especially true when applicants call for justification of the questions in the knowledge test. If a research participants ask "where could I find this piece of information?", the ability to point to the relevant chapter/section becomes a practical measure of validity. This validity is called content validity, and is more enhanced by recognizing or realizing that all questions in the tests were drawn from the information available in the B.O.S.E.T manual (Li, 2003; Kane, 2001; McKnight, 1999; Messick, 1998).

The B.O.S.E.T trainers chose a handful of topics that was considered 'critical' for trainees when working offshore. These topics are:

(a) Emergency response procedures (Basic Safety Induction).
(b) Precaution when moving in smoke (Basic Fire Fighting).
(c) Boarding the life raft (Sea Survival).
(d) Immediate actions and subsequent actions (Sea Survival).
(e) Hypothermia (Sea Survival).
(f) Helicopter rescue and vessel rescue (Sea Survival).
(g) Approaching the helicopter (Helicopter Underwater Egress Training).
(h) Underwater disorientation (Helicopter Underwater Egress Training).
(i) Disembarking (Helicopter Underwater Egress Training).
(j) Life jacket inflation (Helicopter Underwater Egress Training).
(k) Surface abandonment (Helicopter Underwater Egress Training).

Samples of the items in the knowledge test can be presented, as follows:

(1) Once in a life raft, you should immediately:
   a. Go to sleep to conserve energy.
   b. Check for leakage in the life raft.
   c. Check your life jacket for any leakage.
   d. Fire a flare signal.

(2) Imagine yourself on an offshore platform. If you discover a fire on the offshore platform, you should immediately:
   a. Inform the Emergency Response Team (ERT).
   b. Raise the alarm.
   c. Inform your supervisor.
   d. Try to extinguish the fire yourself.

4. Result

4.1. Readability test result

The knowledge test has 14 critical safety questions. The 14 questions are made of 2712 characters, which make up 633 words. There are 152 sentences in this knowledge test. The average number of characters per word is 4.40 characters/word while the average number of words per sentence is 4.16 words/sentences.

Table 3 Readability index of the knowledge test.

Approximate representation of the US grade level needed to comprehend the text	Score
Flesh Kincaid grade level	2.46
Flesch Reading Ease	84.20
Gunning Fog	4.95

The summary of readability tests is shown in Table 3. The Flesch Kincaid Grade Level test value was at 2.46 (2nd Grade – refer Table 1); thus indicating that the average participants aged above 7 or 8 years old should be able to read the questionnaire easily. In addition, the Flesch Reading Ease value was determined at value 84.2 (standard – refer Table 2); hence indicating those above the age of 14 years old should have no problem reading the questionnaire. On top of that, the Gunning Fog Index revealed a value of 4.95; meaning that the knowledge test should be comprehensively understood by the general population. Since the participants were all professionals of above 18 years old, the results suggest that the research participants should not have difficulty understanding and answering the questionnaire (Paul, 2003; Klare, 2000). By these results, it is conclusive that the research tool is readable for the target sample and therefore, the tool have demonstrated Characteristic A (refer Section 2.1).

4.2. Statistical analyses ($N = 79$)

There were 79 Class A offshore professionals that consented being a part of this study, of which 25 trainees were Refresher while the rest were Fresher trainees. The mean test score for 79 participants is 12.20 (SD = 2.31). Knowledge test result from 79 participants yield Cronbachs' Alpha value of 0.76. This result suggests that the knowledge test is reliable for the target sample. The 'discrimination' values for the knowledge test are all above 0.2, which indicates good question discriminations (Trochim, 2005; Mislevy, 2004; Milne, 1999) (Table 4).

For $N = 79$, the knowledge test have again demonstrated that it is highly discriminated (Characteristic C – refer Section 2.1). Table 5 shows the average test scores for both pre and post-training sessions. From the table, the research concludes that the knowledge test is able to distinguish between trained and untrained trainees. Before the B.O.S.E.T training, no participants managed to pass the test (score above the 70%). However, after undergoing B.O.S.E.T training, 65 participants (82.3%) managed to pass the test. This suggests without B.O.S.E.T training, it would have been difficult, if not impossible, for participants to pass the knowledge test.

Table 5 Mean test score (pre and post-training) ($N = 79$).

Items	Pre-training	Post-training
	% correct	% correct
1	22.80	73.40
2	11.40	79.70
3	1.30	86.10
4	13.90	94.90
5	1.30	84.80
6	7.60	79.70
7	20.30	87.30
8	24.10	92.40
9	20.30	89.90
10	10.10	92.40
11	12.70	91.10
12	1.30	92.40
13	6.30	83.50
14	12.70	91.10

From there, the research works to identify whether there is a statistical difference between pre-training and post-training. Fig. 2 shows a histogram of paired sample pre-training and post-training. From Fig. 2, the researcher estimates the data as approximately normal distribution due to the 'bell-shaped' curve. The mean value (10.54) is approximately the same as the median (11) and this implies that the distribution is approximately normal.

Further data analysis using the *P–P* plot test also suggests that the paired data is approximately normal distribution. The result of the *P–P* plot test in Fig. 3 shows that the positions of the plots are situates closely to the diagonal line. This implies an approximately normal distribution. In addition to that, a *Q–Q* plot test was also used as means of determining data distribution (Fig. 4). The plots in Fig. 4 are closely positioned to the diagonal line; hence, there was enough evidence to suggest the paired samples distribution is approximately normal. The research concludes that the paired samples pre-training and post-training is approximately normal distribution.

Table 4 Average score (%) and correlations for each item in knowledge test ($N = 79$).

Items $n(\alpha)$	79 ($\alpha = 0.762$)		Mean M	Standard deviation SD
	% correct	Item-total correlations		
1	73.40	0.38	1.23	0.42
2	79.70	0.27	1.11	0.32
3	86.10	0.35	1.01	0.11
4	94.90	0.26	1.14	0.35
5	84.80	0.57	1.01	0.11
6	79.70	0.56	1.08	0.27
7	87.30	0.30	1.20	0.40
8	92.40	0.39	1.24	0.43
9	89.90	0.40	1.20	0.40
10	92.40	0.32	1.10	0.30
11	91.10	0.46	1.13	0.33
12	92.40	0.32	1.01	0.11
13	83.50	0.40	1.06	0.25
14	91.10	0.31	1.13	0.33

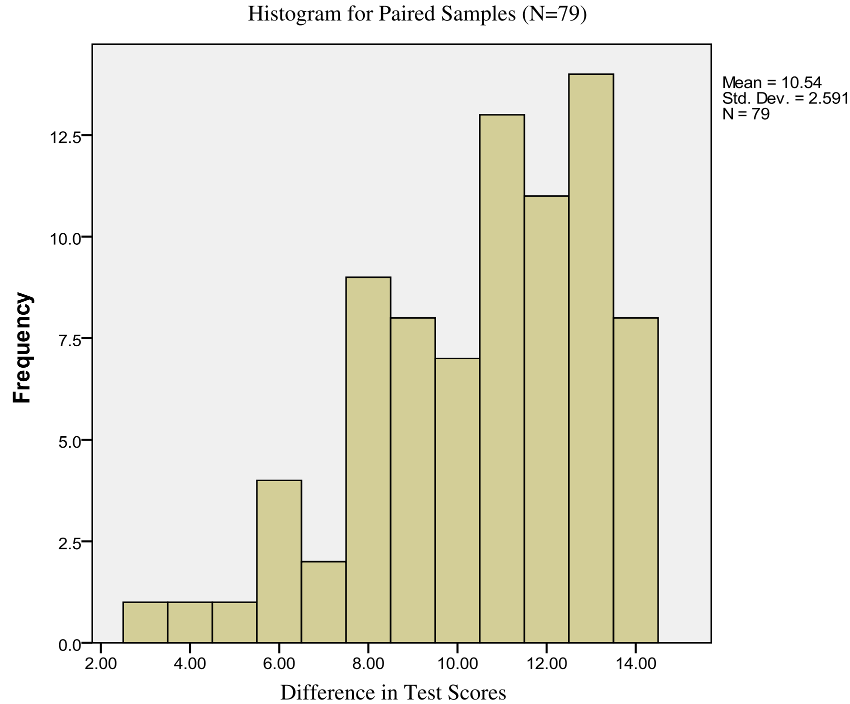

Figure 2 Test scores histogram between pre-training and post-training ($N = 79$).

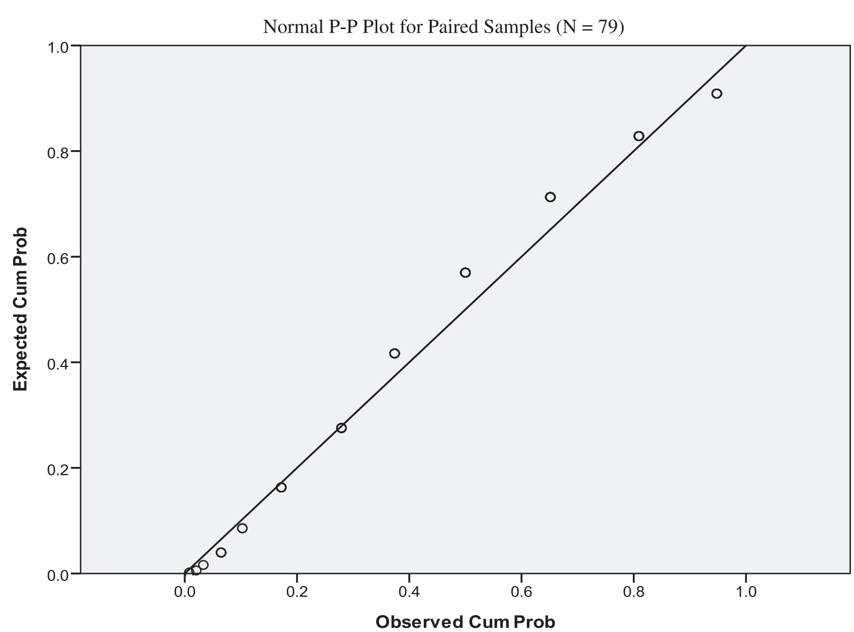

Figure 3 *P–P* plot for paired samples ($N = 79$).

Since the research establishes the paired samples distribution as approximately normal, further analysis requires the research to apply Paired Samples *T* test to identify the *p*-value. The paired samples statistics is shown in Table 6. The mean test score for pre-course was 1.65 (SD = 1.13) while the mean test score post-course was 12.20 ($N = 79$, SD = 2.31).

Table 7 represent the Paired Samples Correlation ($N = 79$). There is not enough evidence to suggest that the paired samples are correlated ($N = 79$, $p > 0.05$). This result is expected as Table 5 illustrates significant difference between the two (2) experimental sessions. Furthermore, the research the average test score difference between pre-course and post-course is 10.54 (SD = 2.59), indicating participants have learned B.O.S.E.T knowledge at a highly significant level, $t(118) = 36.17$. In other words, there is an increase of 75.32% in the average test score. The research is also 95% confident that the mean test score is between 9.96 and 11.12. These statistics are presented in Table 8.

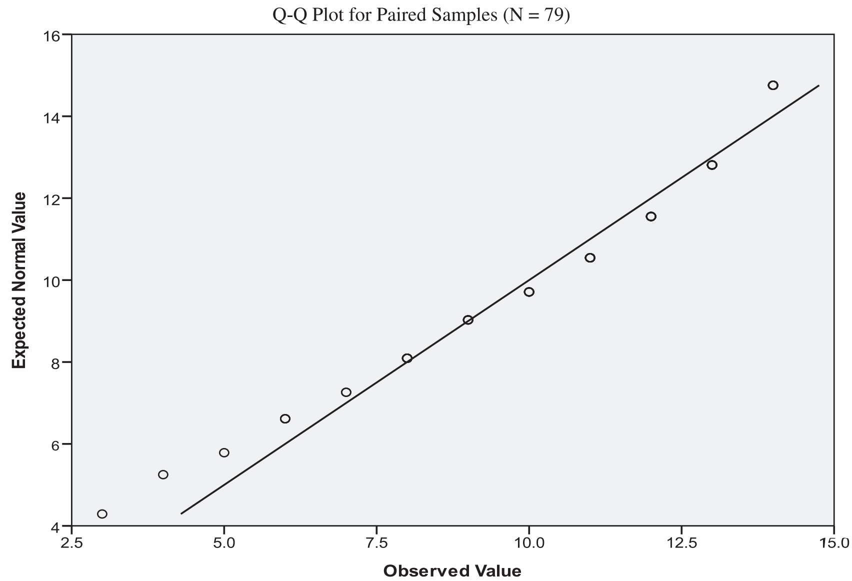

Figure 4 *Q–Q* plot for paired samples ($N = 79$).

Table 6 Paired samples statistics ($N = 79$).

		Mean	*N*	Std. deviation	Std. error mean
Pair 1	Post training	12.2025	79	2.30596	0.25944
	Pre Training	1.6582	79	1.13099	0.12725

Table 7 Paired samples correlations ($N = 79$).

		N	Correlation	Sig.
Pair 1	Post training & Pre training	79	−0.022	0.845

Finally, a Hypothesis Test was used to identify whether there is a significant difference between the two (2) paired samples. The result indicates there is a significant difference at $p < 0.001$. This would suggest that the scores before training and scores after training are statistically different. The summary of the Hypothesis Test is shown in Table 9.

5. Conclusion

The research work to identify the retention rate of critical knowledge among offshore professionals. In order to measure knowledge retention, the knowledge test was developed. Aspects of tool development have been discussed rigorously in this article. Huge amount of efforts have been taken in order to maintain high standards in measuring retention. From reliability test to validity, and finally with readability test, the research tool have proved its credibility. Four hypotheses were used to identify the research tool characteristics. Through mathematical and statistical means, the four hypotheses have been proved correct. Thus, the knowledge test is a valid and

Table 8 Paired samples test result ($N = 79$).

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	After–Before	10.54430	2.59090	.29150	9.96397	11.12463	36.173	78	0.000

Table 9 Hypothesis test summary.

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between before and after equals 0.	Related-Samples Wilcoxon Signed Ranks Test	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

reliable research tool to measure B.O.S.E.T knowledge retention.

6. Impact on the industry

This research will have positive impact in the offshore industry. The fact that the research works to identify the retention capacity of offshore professionals in the field of safety is a very important issue. By identifying the retention capacity, improvements can be made to the course in order to improve knowledge retention. This may encourage safety environments as well as safety habits while working offshore. In an emergency, retention of B.O.S.E.T knowledge can mean the difference between life and death. Thus, by improving the retention rate, this improves chances of survivability in offshore emergency scenarios. The validation of the knowledge test presented in this paper is a critical aspect of the research, as to ensure credibility.

In addition, several early data have suggested support for the Ebbingaus forgetting model research. There are active discussions between the University Aberdeen, Malaysian Oil and Gas companies and CONSIST Sdn Bhd as ways to improve retention among offshore professionals. These include a review on training policies, use of internet based solutions as well as developing an electronic based monitoring system.

Acknowledgements

Special thanks to Mr. Ar Mior Zawari Hassan for the support and collaborations given through CONSIST. CONSIST has always strive to deliver the best quality services to its clients. Since in cooperated in 1996, CONSIST have been receiving local and international recognition for its services. Among the recognition and accreditation received are from PETRONAS to conduct Offshore Training for PETRONAS platform International Maritime Organization (IMO)/Maritime Department of Malaysia to conduct STCW Basic Safety Training for Seafarers, Department of Occupational Safety and Health (DOSH) Malaysia to conduct Certificate in Safety and Health Malaysia and after safety courses, Construction Industry Department Board (CIDB) Malaysia to conduct skill training such as scaffolding and crane, Department of Civil Aviation (DCA) Malaysia to conduct Aviation Safety Training and National Accreditation Board Malaysia to conduct Diploma in Occupational Safety and Health Management. In 2002, CONSIST received ISO 9001-2000 Quality Certificate and in 2004, CONSIST received an excellent award from Malaysia Canada Business Council for best service company category.

In addition, the researcher would like to thank the B.O.S.E.T trainers, Mr. Ruslan Bin Yusof, Mr. Nazri Mokhtar, Mr. Shahril Bin Manaf and Mr. Mohd Fazarul Bin Zakaria for the support and guidance given in this research.

Appendix A. Knowledge test guidelines

Efforts were taken to clearly define the formats of the knowledge test, so as it would not reflect as knowledge survey. These formats include content of knowledge test, alternative answer selections, content of alternatives, position of correct answer, number of alternatives, true–false alternatives, inconsistent alternatives, use of negative forms, high scoring standards, independent questions, and special participants category (McKnight, 1999).

A.1. Content of knowledge test

The knowledge test was designed to measure a sample of critical safety knowledge that offshore professionals should know. All questions developed were referred from the B.O.S.E.T manual, the AGARDograph as well as the help from B.O.S.E.T trainers. To avoid any word confusion, words in the knowledge test were taken before the B.O.S.E.T course manual and the AGARDograph, as a proof that the questions did indeed came from either manuals.

A.2. Alternative answer selections

Each question only has only one correct answer. Efforts were made to ensure that the participants should not have to judge the degree of correctness (for example which is the 'most likely' answer).

A.3. Content of alternatives

A single question was made to address the same piece of information with several plausible alternative answers. The knowledge test works to assess participants' knowledge retention of particular information. If the various choices of answers address different topics, it would be impossible to determine from the responses what knowledge it is that participants do or do not retain.

A.4. Position of correct answer

The position of the correct answer in the series of alternative answers was at random, in order to prevent participants benefiting from systematic patterns (McKnight, 1999).

A.5. Number of alternatives

To reduce chances of guessing the correct answer, the numbers of alternative answers was set to four. A standard of four alternative answers to one question was used in the knowledge test.

A.6. True–false alternatives

The knowledge test was designed to avoid true–false answer because of the relatively high opportunity of guessing the correct answer (McKnight, 1999).

A.7. Inconsistent alternatives

Efforts were made to avoid sentence inconsistencies that attract attention to a particular alternative answer. These include alternative answers being longer than others, use of attractive words or jargons, and rationalizing incorrect answers to make them appear plausible (McKnight, 1999).

A.8. Use of negative form

The knowledge test was designed to avoid negative question forms such as "Which is not...". This type of questions requires participants to search for an incorrect answer. There is a chance that this would create an opportunity for forgetful participants to choose wrongly (McKnight, 1999).

A.9. High scoring standards

A 70% score was determined to be the passing mark for the knowledge test, due to huge amount of information that participants require to retain after training. The 70% passing mark has been widely used in many knowledge test standards such as the United States Department of Aviation knowledge test guide (Yang, 2008).

A.10. Independent questions

The questions in the knowledge test were designed to be different from one to another. This was done with the intention of minimizing opportunity of hints for other questions. In addition to that, the knowledge test was designed to avoid any two questions in the same form to address single information.

A.11. Special participant category

The B.O.S.E.T knowledge test was conducted in Malaysia. It is not uncommon to have participants that were not able to speak or read in English, since Bahasa Malaysia is the mother tongue language. However, the inability to read or speak

Job class accepted in the research

Occupation	Accepted (Yes/No)	Occupation	Accepted (Yes/No)
Electrical Engineer	Y	IT Engineer	Y
Welder	N	Coring Engineer	Y
Production Manager	Y	Pipeline Engineer	Y
Field Technician	N	Instrument Drafter	N
Engineer	Y	Project Coordinator	Y
Project Manager	Y	Draughtsman	N
Area Engineer	Y	Structural Engineer	Y
Drill Supervisor	Y	Operations Engineer	Y
Piping Designer	Y	Safety Consultant	Y
Mechanical Engineer	Y	EIA Engineer	Y
Drilling Engineer	Y	MCI Engineer	Y
Technician	N	Geoscientist	Y
Process Engineer	Y	Surveyor	Y
Rig Manager	Y	Crane Operator	N
Project Engineer	Y	Deck Officer	N
Sales Engineer	Y	Subsea Engineer	Y
Barge Master	N	Instrument Engineer	Y
Service Engineer	Y	MWD Engineer	Y
Manager	Y	Corrosion Engineer	Y
Roustabout	N	Mud Engineer	Y
Superintendent	Y	Technical Assistant	N
Safety Officer	Y	System Operator	Y
Instructor	N	Rig Accountant	N
Reservoir Engineer	Y	HSE Assurance	N
Trainee Engineer	Y	Security Auditor	N
Petro physicist	Y	CAD Designer	Y
Pump man	N	Chain Manager	Y
Seaman	N	Process Manager	Y
Marine Engineer	Y	Service Manager	Y
System Analyst	Y	Drilling Superintendent	Y
Medical Officer	Y	Field Engineer	Y
Programmer	Y	Offshore Installation Manager	Y
Diver	N	Procurement Specialist	N
Maintenance Engineer	Y	General Manager	Y
Technical Executive	N	Instrumentation Engineer	Y
Chief Officer	N	Planner	N
Support Engineer	Y	Safety Coordinator	Y
Cook	N	Oiler	N
Director	N	Field Supervisor	Y
Installation Engineer	Y	H2S Engineer	Y
Inspection Engineer	Y	Mooring Master	N
Assistant Manager	Y	Service Technician	N
Costumer Representative	N	DWM Engineer	Y
Project Consultant	Y	Rig Engineer	Y

English was not a stumbling block for this research. This problem was solved by developing the knowledge test in the Bahasa Malaysia language, as an alternative for those who do not speak or read English.

References

Allen, M.J., Yen, W.M., 2002. Introduction to Measurement Theory. Waveland Press, Long Grove, IL.

Berden, H.J., Willems, F.F., et al., 1993. How frequently should basic cardiopulmonary resuscitation training be repeated to maintain adequate skills? BMJ 306 (6892), 1576–1577.

Bradburn, N., Sudman, S., Wansink, B., 2004. Asking Questions. Jossey-Bass, San Francisco, CA.

Collins, E., 2000. As good as new? Occupational Health July, 21–24.

Cullen, M.C., 1992. First aid retention of knowledge survey. HSE Research Paper, 32. ISBN 01-886380-0.

Curry, L., Gass, D., 1983. Effects of training in cardiopulmonary resuscitation on competence and patient outcome. CMAJ 137, 491–496.

Darville, R., Hiebert, M., 1985. Small Claims Court Materials: Can They Be Read? Can They Be Understood, Canadian Law Information Council, No. 7.

Deliere, H.M., Schneider, L.E., 1980. A study of cardiopulmonary resuscitation technical skill retention among trained EMTs. EMT Journal (2), 57–60.

DuBay, W.H., 2006. Smart language: Readers, Readability, and the Grading of Text. Costa Mesa: Impact Information.

Fossel, M., Kiskaddonm, R.T., Sternbach, G.L., 1983. Retention of cardiopulmonary resuscitation skills by medical students. Journal of Medical Education 58 (7), 568–575.

Fry, E.B., 2002. Readability versus leveling. Reading Teacher 56 (3), 286–292.

Hargis, G., 2000. Readability and computer documentation. ACM Journal of Computer Documentation 24 (3), 122–131.

Harvey, H.D., Fleming, P., 2003. The readability and audience acceptance of printed health promotion materials used by environmental health departments. Journal of Environmental Health 65 (6), 22–29.

Hubbard, D.W., 2009. The Failure of Risk Management. Why it's Broken and How to Fix it? John Wiley & Sons, USA.

Kane, M., 2001. Current concerns in validity theory. Journal of Educational Measurement 38, 319–342.

Klare, G.R., 2000. Readable computer documentation. ACM Journal of Computer Documentation 24 (3), 148–168.

Kumar, R., 2005. Research Methodology, second ed. Pearson Education, Singapore.

Lacity, M., Jansen, M.A., 1994. Understanding qualitative data: a framework of text analysis methods. Journal Management Information System 11, 137–160.

Li, H., 2003. The resolution of some paradoxes related to reliability and validity. Journal of Educational and Behavioral Statistics 28, 89–95.

McKnight, A.J., 1999. Guidelines for Knowledge and Skills Testing. American Association of Motor Vehicle Administrator.

Messick, S., 1998. Test validity: a matter of consequence. Social Indicators Research 45, 35–44.

Milne, J., University of Aberdeen, 1999. <http://www.icbl.hw.ac.uk/ltdi/cookbook/info_questionnaires/index.html> (accessed 07.07.10).

Mior, A.Z., Ramanie, H., 2009. CONSIST Offshore Safety Training, Basic Offshore Safety and Emergency Training (B.O.S.E.T) manual, revised ed. Kuala Lumpur.

Mislevy, R., 2004. Can there be reliability without reliability? Journal of Educational and Behavioral Statistics 29, 241–244.

Ogince, M., Hall, T., Robinson, K., Blackmore, A., 2007. The diagnostic validity of the cervical flexion-rotation test in C1/2-related cervicogenic headaches. Manual Therapy 12 (3), 256–262.

Paul, T., 2003. Guided Independent Reading. School Renaissance Institute, Madison, WI, http://www.renlearn.com/GIRP2008.pdf.

Redish, J.C., 2000. Readability formulas have even more limitations than Klare discusses. ACM Journal of Computer Documentation 24 (3), 132–137.

Rosenthal, R., 1998. Covert communication in classrooms, clinics, and courtrooms. Eye on Psi Chi 3 (1), 18–22.

Schmitt, N., Landy, F.J., 1993. The concept of validity. In Schmitt, N. & Borman, W. C. (Eds.), San Francisco.

Schwarz, N., Oyserman, D., 2001. Asking questions about behavior: cognition, communication, and questionnaire construction. American Journal of Evaluation 22 (2), 127–160.

Shuttleworth, Martyn, 2009. Face Validity. Retrieved [Date of Retrieval] from Experiment Resources. <http://www.experiment-resources.com/face-validity.html>.

Trochim, W., 2005. Research Methods: The Concise Knowledge Base. Atomic Dog Publishers, Cincinnati.

Trochim, W., Donnelly, J.P., 2007. The Research Methods Knowledge Base, third ed. Thomson Publishing, Mason.

Wegner, M., Girasek, D., 2003. How readable are child safety seat installation instructions? Pediatrics 111, 588–591.

Wells, C.S., Wallock, J.A., 2003. Instructor Guide to Understanding Test Reliability. University of Wisconsin, USA.

Zakaluk, B.L., Samuels, S.J., 1988. Readability: It's Past, Present & Future. International Reading Association, Newark, Delaware.