Competence Assessment Using Simulators and Automated Assessment Tools

Alan Whitcher
Warsash Maritime Academy, a faculty of
Southampton Solent University
UK

Ray Gillett
MPRI, a division of L-3 Services, Inc.
UK

Gastech 2009
Abu Dhabi National Exhibition Centre
UAE

25 – 28 May 2009

Abstract

With the shortage of qualified seafarers the issue of finding competent personnel to man both the new and existing generations of LNG and LPG Carriers is becoming a major issue. Whilst the certification system provides a means of verifying if a seafarer has the appropriate knowledge and understanding, verification of their ability to undertake a task or 'competence' is more difficult especially away from the actual physical environment. The use of simulators provides an environment where this can be done safely but the results have necessarily been subjective, based on individual instructors. New tools are now becoming available which attempt to take the subjectivity out of this type of assessment and also make the process easier to complete. The purpose of this paper is to explain the current position and show how at least one of these tools may be used.

This paper begins by examining the difference between assessing underpinning knowledge and the higher level cognitive processes commonly referred to as 'competence'. A number of competence assessment methodologies are described and compared with a comparison of their relative advantages and disadvantages including the role of the assessor and the limitations of assessment by individuals. The paper then explains how assessment can be undertaken by automated means including a comparison with some of the systems that are currently available. A detailed description of how the MPRI automated assessment tool can be used is provided. The description includes details of the tests undertaken, test conditions and the subsequent results. Finally future development and testing of this system is described together with a description of how this might be used in the assessment of competence within the maritime community.

Competence Assessment Using Simulators and Automated Assessment Tools

Defining Underpinning Knowledge and Competence

Underpinning knowledge may be considered as the bedrock of information required before a task can be conducted successfully. This knowledge may consist of domain knowledge and procedural knowledge. In the context of gas ship operations, domain knowledge may relate to the particular vessel, its cargo system and equipment and regulations set out in the IGC code. Procedural knowledge may relate to operating procedures set out by the company, in the vessel's operating manual and guidance in the Tanker Safety Guide (Liquefied Gas).

Cognitive learning is about the acquisition and use of knowledge and in 1956 Bloom compiled a taxonomy of learning which is widely recognised today. Bloom's taxonomy comprised six increasingly complex and abstract levels within the cognitive domain. In ascending order, the levels within Bloom's taxonomy are;

- Knowledge
- Comprehension
- Application
- Analysis
- Synthesis
- Evaluation

Knowledge is the lowest level of these levels and some verbs which represent intellectual activity at this level are; define, duplicate, list, memorise, name, recall, recognise, repeat, state. These verbs are often used in multiple choice questions which are used in assessments at training centres. Those who pass such assessments are often declared competent and given certificates to attest this. However this lowest level of knowledge is a long way from what most ship operators are looking for when they talk about competent operators.

Wolf supports the view that fragmented knowledge 'facts' alone do not equate with competence " A large amount of blood has been shed, in the school as well as the vocational sector, over the question of whether or not there is value in the learning of 'bucketfuls of facts'. The debate is not unrelated to the equation of 'knowledge testing' with measurement of factual recall, often in the most fragmented kind – and with the prevalence of just such tests in the vocational sector. However, what we should be talking about when discussing the relationship of knowledge to competencies is something much wider than factual recall. It is what much of the literature on medical competence refers to as 'cognitive abilities'." (Wolf 1989)

It might be argued that a competent person is someone who can demonstrate all of the levels of learning listed in Bloom's taxonomy, however all of these relate only to the cognitive domain. In 1975 Dave produced a similar taxonomy for the psychomotor domain, in ascending order;

- Imitation – observes skill and tries to repeat it
- Manipulation – performs skill according to instruction rather than observation
- Precision – reproduces a skill with accuracy, proportion and exactness
- Articulation – combines one or more skills in sequence with harmony and consistency
- Naturalisation – completes one or more skills with ease and becomes automatic

(Dave 1975).

Smith and Hancock provide a definition of competence which combines all of the above; "Competence melds knowledge, abilities and cognition to generate appropriate behaviour given conditions in the task environment. Competence resides in the agent and enables skilled performance" They add, …"the behaviour that competence generates is the agent's solution to the problem of knowing those cues and demands in the environment that enable it to take action that aligns with the current goals" (Smith and Hancock 1995).

Clearly a competent operator is someone who can do more than recite a handful of facts. A competent operator must be able to take domain knowledge and demonstrate the ability to synthesise this with the current status of the plant and apply mental models to be able to achieve current goals. Underpinning knowledge is just the first part of this process.

Measuring Underpinning Knowledge and Competence

Underpinning Knowledge

It is not surprising that since knowledge is easily defined by verbs such as; define, duplicate, list, memorise, name, recall, recognise, repeat, state etc. that questions incorporating these verbs have become the most common method of assessment in maritime training centres. For ease of application assessments asking for such demonstrations of knowledge are often in the multiple choice format. However, multiple choice questions require the construction of realistic distracters, without which, the correct answer may be easily identified. Multiple choice questions also mean there is no requirement to formulate an answer, as would be the case with an open question, which would demonstrate a better depth of knowledge and understanding. When assessing mariners, the fact that the language used to write the question may not be the first language of the candidate, is sometimes cited as reason not to use open questions, removing the possibility that the test itself becomes one of language not subject knowledge.

Negative marking can be applied to multiple choice questions to limit the guessing of answers. In this case a wrong answer would result in the subtraction of one (or more) marks. The idea of this approach is that if the candidate is unsure of the answer it is safer not to answer at all rather than have a guess whilst in a traditional marking scheme without negative marking, he has everything to gain and nothing to lose. However this approach does not appear to be commonly used in maritime training centres.

Open questions can be more insightful to the candidate's underpinning knowledge and can be used to demonstrate the higher levels of knowledge in Blooms taxonomy. However there is an additional burden on the assessor to read and interpret the answers against a clear marking scheme. There is also the additional burden on the candidate to articulate written answers in a language which may not be their mother tongue.

Multiple choice and open questions are a reasonable way to assess the lower levels of underpinning knowledge. Questions of the open type are arguably better at demonstrating some of the mid levels of learning such as compression, application and analysis. An alternative way to administer open questions is orally. This reduces problems where the language used is not the first language of the candidate but it means assessment is on a one to one basis. Oral questions raise the issue of recording evidence of questions asked and answers given in order to ensure that assessments are objective, auditable and transparent.

Competence

When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind

William Thomson, Lord Kelvin (1824 - 1907)

Competence is a construct and as such is difficult to measure. As can be seen from the definitions above it is more than just the completion of a task. While it starts with knowledge it incorporates other things and other constructs such as mental models and situational awareness. As with most constructs there is much debate about definitions and even more about ways to measure them. The less clear the definition, the less likely it will be that a good measure can be made.

Knowledge and understanding is not divorced from competence rather they contribute to it and may be inferred from observable behaviour. "In developing competence based vocational training and qualifications in this country (the UK), most of the emphasis has been on 'output' measures. It has also been policy to encourage the direct assessment of 'performance', ideally in the workplace, as the preferred measure of competence." (Wolf 1989).

Liquid cargo handling simulators provide the opportunity for candidates to demonstrate their competence in conducting cargo related operations in a safe environment. However, to measure a candidate’s performance to determine if they are competent in undertaking those tasks is not a straightforward process. There needs to be a clear and consistent standard to measure against and a mechanism to measure with. Without this an assessor who simply watches candidate performance is only providing a subjective opinion which will vary between assessors.

Vocational training qualifications typically identify performance as the best possible measure of competence. Vocational qualifications are comprised of units which are made up of elements. Each element is then broken into performance criteria. Range statements may then be applied to individual elements and performance criteria to clarify the conditions under which they must be conducted.

In the UK the National Council for Vocational Qualifications (NCVQ) defines competence as the ‘ability to perform an activity within an occupation’. In order to produce the units, elements, performance criteria and range statements for a vocational qualification a functional analysis of the activity is required. In the context of LNG ship operations the SIGTTO LNG Shipping Suggested Competency Standard provides a good starting point for this. Although written as a competence based standard i.e. one in which statements refer to outcomes rather than activities, it doesn’t provide the same level of detail as a Vocational Qualification (VQ).

In the context of using simulation to demonstrate competence the functional analysis of operations and the identification of units, elements, performance criteria and range statements to generate an objective assessment within a simulator has been used at Warsash Maritime Academy for a number of years for assessing LNG and FPSO cargo control room operators and has been shown to be successful.

Currently at Warsash, when the simulator is used for assessment, the process is started by developing a consistent ‘standard’ or base line against which an assessment can be made. All the tasks that must be achieved are broken down into component parts by a functional analysis resulting in a structure which is similar to the vocational standards described above. A unit might be ‘Commence Loading’, which could be broken down into elements and performance criteria thus;

Unit 4 Commence Loading Cargo Tanks
Element 1 – Line up to Tanks

Performance criteria

- Keeps manifold ESD valve closed
- Opens valves from manifold to liquid header
- Opens tank branch valves
- Opens tank filling valves

Element 2 – Prepare High Duty Compressor

Performance criteria

- Sets Nitrogen pressure to compressor seals
- Lines up lube oil system
- Starts lube oil cooling pump
- Starts auxiliary lube oil pump
- Ensures High Duty Compressor alarms are off

Element 3 – Line up High Duty Compressor to return gas to shore

Performance criteria

- Etc., etc.

The list of performance criteria are converted into a checklist which is ticked off by the assessor as he observes the candidates actions. This approach provides a clear and repeatable standard against which all candidates are measured.

During the observation of performance, an assessor would not only note the correct actions on a checklist but would also identify incorrect actions that were carried out. When observed, incorrect actions would be noted for discussion with the candidate. At the end of each exercise there is the opportunity for an oral interview during which the candidate has the chance to justify actions taken and the assessor may question incorrect actions or other observations. This adds further insight into the depth of knowledge and understanding of the candidate and assists in the determination of competence.

By using this methodology the assessment becomes objective and standardised since all assessors must follow the same standard. Vocational qualifications use other forms of evidence in addition to practical demonstration, which normally results in a candidate accumulating a portfolio of evidence to be assessed. Whether or not that would be of value in the domain of gas ship operations is a separate discussion.

The main limitation of this method of assessment is that it requires a one to one assessment. It is not possible for a single assessor to observe every action of more than one candidate. An automated assessment tool which could observe more than one candidate at a time, and record the completion of all the performance criteria in the same way a human assessor would, but for multiple candidates, would clearly be desirable.

Automated Assessment Methods

Automated assessments have been used in simulation for a number of years, however the methodologies used to measure, and where applicable determine competence, have been limited.

The reasons for this primarily relate to technology but also the purpose for which they are being used. Traditionally assessment of a candidate on a simulator has been made on a one to one basis and consequently the assessment system only has to provide sufficient information or evidence to assist the instructor in making his final decision. However when assessing multiple candidates at the same time (as can be the case when multi station simulators are being used) the information provided by these systems is insufficient to deal with the complexity.

For the purposes of this paper a comparison is made with the assessment system used within the MPRI driving simulators. These simulators have a range of driving scenarios which can be selected each including a predefined set of rules that look for errors made by the candidate. If the candidate makes one of the identified errors the event is identified and reported. For each error a specific score is assigned. However, one of the disadvantages is that the errors are not context specific, meaning that whenever they take place the penalty score is always the same irrespective of the speed or situation the vehicle is in.

At the start of an exercise the overall assessment score is set at 100.Every time the candidate makes an error the value is deducted from the initial 100 total resulting in a final score. A report is produced at the end of the exercise listing all the errors, and the final score. This system does not indicate a ‘pass’ or ‘fail’ it simply records a score which the instructor can then use to assist in their final judgement and also provide documentary evidence to support the overall assessment of the candidates competence during that exercise.

To enable a system such as the one described above to be used for assessments of multiple candidates at different times by different instructors it is important that the identified errors used are always consistent and are agreeable by the majority of the industry involved. Because experience in driving is widespread, the errors used within the MPRI system are based on the results of detailed research carried out over a number of years by the leading driving organisations. This ‘validation’ to a degree offsets the limited scope that the number of errors that can be checked for.

In industries where such detailed research is not available (such as marine operations) obtaining a definitive set of events to be checked for is very difficult, as many instructors will interpret situations differently depending upon their own previous experience. This problem is further compounded when error checking cannot include the context at the time the error occurred, as in the system described here.

Automated Assessment Systems in the Maritime Industry.

With the decrease in available experienced manpower in the maritime industry over the last decade the problem of finding people who can become skilled instructors and assessors has become much more difficult. At the same time the requirement for documentation to support certificates issued has increased due to the various audit requirements that are now commonplace. As a consequence training establishments have been asking for more assessment tools to be provided along with new training equipment, such as simulators, to ease the burden that they now experience. Consequently the manufacturers of the equipment have responded by providing more tools.

A number are based on the use of 'logic gates' which allow an instructor to build their own rules by combining true/false statements and then assign a score which they think is appropriate. Whilst providing the basic tool set the problem with such systems is that the final result is limited by;

a) The ability to build complex patterns using the limited true/false functionality provided by the tools
b) The need to have a large number of rules checking the possible assessment criteria in order to provide a good assessment. (Identifying the criteria that have to be included to ensure the required result is achieved from a test is a complex task and many instructors may not have the time or experience to complete such a task successfully)
c) The lack of published research on the criteria that should be used. Subjectivity of the final result will always be an issue especially when only simple rule formats are used.
d) How well the instructors can devise an appropriate scoring system so that the effect of each event identified has a balanced effect on the overall assessment

MPRI Automatic Assessment Tool

The ideal automated assessment tool would:-

a) allow assessments to be conducted on multiple candidates / simulators at the same time
b) be one which could have all of the knowledge and experience of a human assessor who can identify both correct and incorrect actions
c) be without the subjective views which arise to some degree when a person makes a judgement about an observed performance.
d) not be subject to the distraction that a human assessor may experience which could also influence the judgement they derive from observed performance.
e) Be able to automate the scoring process.

To achieve this MPRI developed a system based on the use of an 'Expert System', initially developed to provide a replacement for an instructor in a teaching role, but which has now been expanded to include detailed assessment.

An Expert System is a form of Artificial Intelligence that is based on the use of rules which when combined together can define 'patterns' of results and behaviour that need to be identified. The reason that an Expert System was used as opposed to some of the more popular areas of AI such as Neural Networks and Fuzzy Logic, is that by using rules which test for specific conditions the expected results are always known and hence can be tested and verified. With Neural Networks, Fuzzy Logic and similar technologies, the expected output may change as the systems 'learn' as more inputs are taken, consequently it is very difficult to produce a system that can be proved to behave as expected.

An Expert System allows an expert to input rules that the system then combines in a similar way as the human brain might be considered to work, i.e. taking in multiple inputs, and processing them to identify any patterns that may be in a previously defined database, and if so identify a specific event or predict outcomes on what has gone before. Consequently it can be set up to be capable of identifying correct and incorrect actions in the same way as a human assessor, and importantly, it is able to determine context in a similar way to a person, so it can determine if a particular action is correct or not in a given context.

For example inert gas can be delivered to the top or bottom of a tank. In one context delivery to the bottom of the tank is correct (such as primary inerting after drydock), but in another, delivery to the top of the tank is correct (such as purging a tank containing Propane vapour prior to gas freeing). A human assessor will recognise the context and automatically apply the correct rule. In a simulator a simple rule, such as that found in the driving assessment simulator will not work when it is context dependent. However by adjusting the rules within the expert system the context can be easily recognised and the rule dealt with accordingly.

The MPRI automatic assessment tool starts by capturing all the patterns of actions which are acceptable and unacceptable from a subject matter expert (SME). This is then translated into rules and tests which are processed by the Expert system.

An average ship model has approximately 3000 variables which can be used to create rules and patterns of rules. The raw measurements can be taken over time to provide static, velocity, acceleration and surge differential values. Any number of concepts can be either derived through predicates (logic statements in the rules) from variable and scores, or from switches (truths) via patterns, holistic or Gestalt patterns, or patterns of probability (Bayes). This provides a mechanism to assess patterns of concepts at any level of abstraction. (Addis and Gillett 2007, p9). To provide the ability to detect complex patterns any rules produced can be nested together where required.

When the initial rules used to detect the context or specific event have been created a 'test' is produced using these rules. For example, if the operation of Loading an LNG Carrier is considered, a series of rules may initially be created about how the HD compressor should be set up, started and operated. Additionally rules defining incorrect actions may also be defined. These rules are now combined into single or multiple tests. Each test will identify both correct and incorrect actions relating to the operation of the compressor.

Further rules and tests for other parts of the loading operation will be constructed such as for ballasting, lining up of cargo tanks, management of tank pressure, management of tank levels, amount of trim and list. All of these tests are then joined in one overarching test for the entire operation of loading (see fig 1). Thus many possible patterns of actions by the candidate are possible, all of which are analysed by the software resulting in an overall pass mark for the operation. The process by which this is achieved is illustrated below.

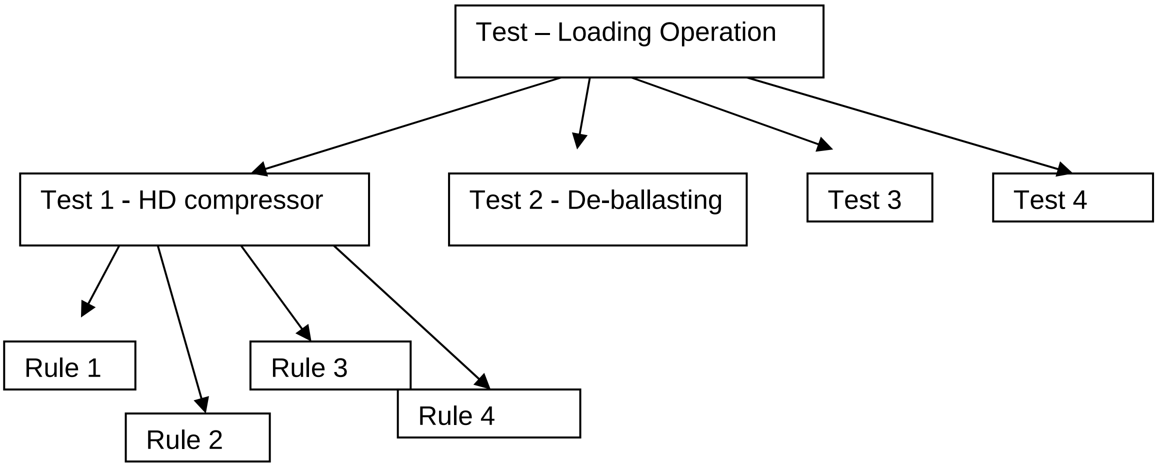

Fig 1 Rule and Test Relationships

Marking the Assessment

Identifying correct and incorrect actions is only the first part of the assessment process. Assigning marks to the actions is also required in an automated system to determine success or failure. It is not normally practical for a human assessor to manually assign marks to actions while observing performance in real time, and if there is no defined standard to measure against the final decision as to whether the candidate is competent is a subjective one. Clearly if the candidate does everything correctly he has passed. If he has done everything incorrectly he has failed. It is more likely that some errors will have been observed and it is these which must be judged by the assessor. Was the error one of fundamental principle, which would result in failure, or was it of a more minor nature which might be acceptable? How many minor errors are acceptable?

The 'WISE' Automated Assessment software includes an automatic mark assignment mechanism which produces an optimised marking scheme. It also includes a confidence assessment with the marking scheme so the chances of passing a novice candidate and failing a knowledgeable one are minimised.

Since the expert system can identify actions and patterns of actions it can assign points (both positive and negative) to those actions. As a human assessor we would judge a situation not by individual actions alone, but by patterns of actions collectively. We don't mentally assign a point each time a valve is opened or closed as part of setting plant status, but we recognise that the plant is in the correct status by the pattern of actions that have been taken. Some patterns are desirable, (correct actions), while some are undesirable, (incorrect actions).

The WISE Automated Assessment Tool analyses all possible pattern combinations utilising the Bayesian Theory of Classification and Bayes' rule (Addis and Gillett 2007). The result is a series of patterns that represent failures and a series of patterns which represent passes. Some pass patterns are more correct than others and equally some fail patterns more incorrect than others. At the extremes the candidate has clearly passed or failed, but there is a zone of overlap between pass and fail patterns where there is the risk of passing an ignorant candidate or failing a knowledgeable one. This can be likened to the task faced by the human assessor who has to make a decision on a candidate who has made some errors and is therefore not a perfect pass or an obvious failure.

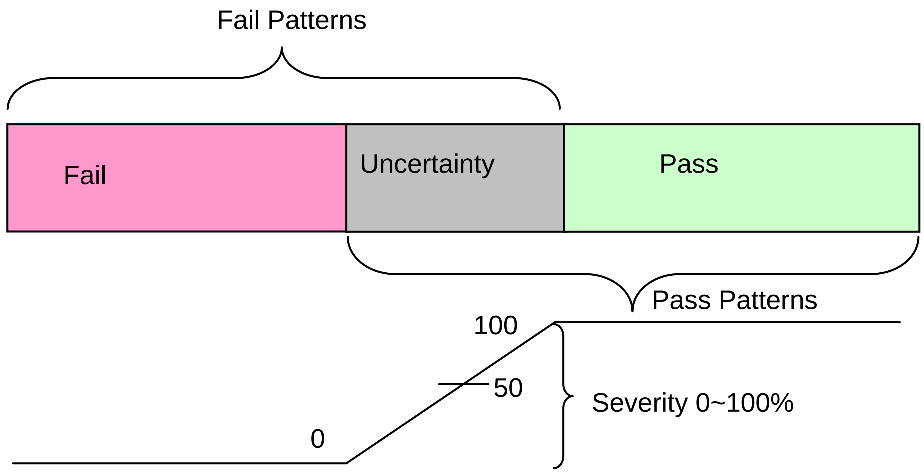

Fig 2 An Estimate of the probability of Pass or Fail (from Addis and Gillett 2007)

When a test is constructed a severity option is available to select from 0 to 100%. 0% severity puts the pass mark at the lower end of the grey uncertainty range (see fig 2), meaning that it is possible that a candidate demonstrating a fail pattern will pass. 100% puts it at the top end of the grey uncertainty range, meaning that it is possible that a candidate demonstrating a pass pattern could fail, however only a candidate who demonstrates clear pass patterns can pass. Thus the degree of confidence in the result can be adjusted to suit the importance of the test.

The responsiveness of the test results can be adjusted and will affect how quickly the overall score for test will be affected. The more critical the test the more responsive it may be, so for example if a negative action was very critical a short response time would see the negative score incurred applied immediately, while a less critical action might affect the score gradually over time. Tests may be nested as required.

Monitoring Student Progress

The WISE Assessment system monitors the status of the vessel and all the operations that are being undertaken within the simulation exercise, in real time, and provides the current candidate score for each test every 15-20secs by use of the methods described below. The progress of each Candidate can therefore be monitored throughout, rather than just providing a score at the end of the operation. This provides the assessor with an overview at a glance of progress at any given stage. Progress is demonstrated in two ways.

1. Tree Structure of Tests and Rules.

As can be seen from the fig 3 the overarching test is expanded into its component tests, which are in turn expanded into their component rules. Tests and rules have a unique numerical identifier which is prefixed with the letter 'T' for a test and 'R' for a rule.

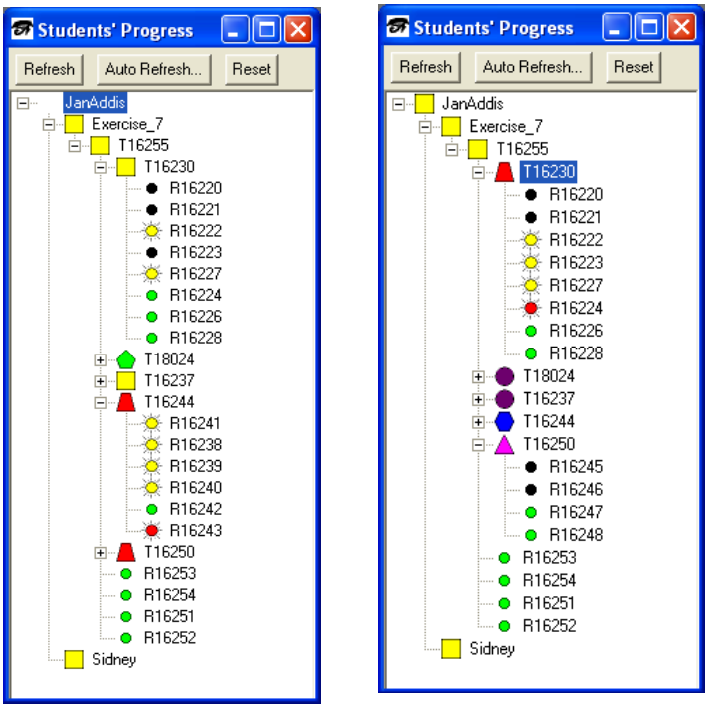

Fig 3 Examples of student progress by Tree Structure diagram

Some rules are positive, (things which should happen) and these are indicated with a black circle to indicate that it 'should happen but has not yet', while a yellow circle indicates that it 'should happen and actually has'.

Some rules are negative, (things which should not happen). A green circle indicates it shouldn't happen and hasn't happened, while a red circle indicates it shouldn't happen but has.

Clicking on any rule number brings up the rule in a separate window so the assessor can identify what the rule is at any time and therefore what the required action was (for example R12345 - Glycol pump inlet and outlet valves are open, the pump is running and the discharge pressure is between 20 and 200 kPa).

Performance within tests is broken into levels which are indicated by symbols. There are 7 different levels depicted as a red triangle indicating a complete failure through to a purple circle indicating the perfect student. (Each incremental symbol has an additional side to its perimeter). All of these symbols are updated automatically at user defined intervals, (for example every minute), which provides a continuous dynamic indication of candidate performance.

ICON	Rules
	Correct Action – Not yet carried out
	Correct Action – Carried out
	Incorrect Action – Not yet carried out
	Incorrect Action – Carried out

ICON	Score	Meaning
	At or above 99%	Excellent
	Above 100% severity	Definite pass
	Above or at pass mark	Pass
	Below pass mark and above 50% severity	Possible fail
	Below pass mark and below 50% severity	Fail
	Below 0% severity	Definite fail
	In fail pattern zone	No chance

Fig 4 Status Indicators Used in the Tree Structure Diagram

2. Student Progress Graph

The second method of observing and recording student progress is in a graphical format as shown in fig 5. This displays the current score for the overarching test for the operation and the sub tests from which it is comprised. As more individual tests are completed satisfactorily so the overall test score rises. Equally if things go wrong then the individual test and overall test scores will fall. The response value used when the test was created determines how quickly positive or negative scores act on the graph.

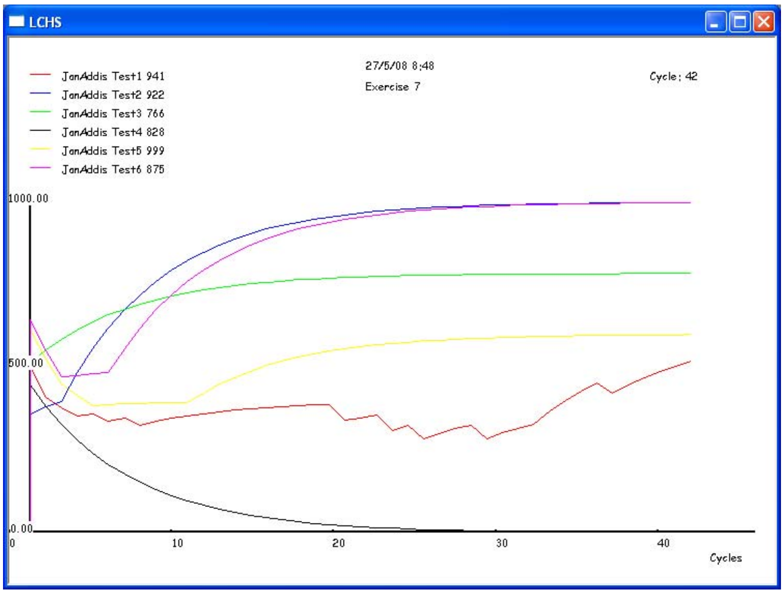

Fig 5 Plot of scores

3. Final Report

On completion of the assessment the automated assessment tool will produce a final report for each candidate. The report will summarise the following;

- Overall score
- Score for each component test
- Rules correctly activated for each test (correct actions)
- Rules incorrectly activated for each test (incorrect actions)
- Graphical summary of test scores throughout the operation

This detailed record will show where errors were identified which can be used to recommend further training which may be required. The complete record can then be retained for later audit.

Testing of the WISE Automated Assessment Tool

The automated assessment tool has been used on a 6 student station MPRI simulator at Warsash Maritime Academy. It has been shown that it can successfully monitor and record the actions of all six students simultaneously during training courses. The tree structure of tests and rules shows at a glance how students are progressing and allows for easy identification of actions that have not been taken or of incorrect actions as can be seen in Figure 6.

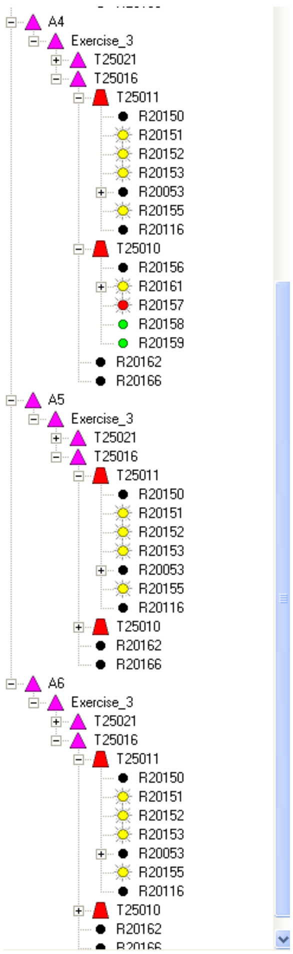

Fig 6 Student Progress During an Exercise

In a training scenario this allows the instructor to quickly identify which student may need additional support. The graph view shows the progress achieved through the sub tests that comprise the exercise, although this does not give the detail of actions not taken or incorrect actions that are possible with the tree view. The graph view can provide a graphical view for each student in a tiled format, which will provide an overview of all students at a glance with the ability for any one of these graphs to be capable of being selected and expanded to provide a detailed view as required.

Addressing the needs of the Instructors and the Maritime Industry

During the development of the automated assessment system it became very evident that whilst providing the tool could assist in allowing instructors with their assessment requirements, because of the time and detail that is required to identify the correct rules to enter to produce the results required, only a few users would be likely to explore the system's full potential. To address this issue and to allow the system to be used 'off the shelf' MPRI will be supplying predefined assessment databases for the respective ship models as required, based on industry standard training requirements. The benefits of this approach are that it:

- will allow the instructors to use the assessment gathering capabilities without the need to allocate a lot of preparation time before being able to use the system

- allow assessments undertaken in different institutions and by different assessors to be directly compared, so assisting to remove the subjectivity currently present in the industry

Conclusions

No automated assessment system in a simulator can be the final arbiter of competence. Ultimately that decision is one that is made by a human assessor. However, the automated assessment system described provides support to the assessor in a way which is objective, repeatable, recordable and able to do so with multiple students concurrently. In some cases the system will indicate a clear pass or fail, but where there is some doubt as to a pass or fail, errors made are clearly identified which may allow the assessor to question the candidate in those areas before finally reaching a decision. A competence assessment is unlikely to be conducted over a single operational exercise, so the assessor will have documented evidence from a number of assessment exercises which he can consider before reaching a final decision.

Although testing so far has only been in the training scenarios rather than formal assessments, it is clear that this system provides useful summative feedback to the instructor/assessor in a clear, graphical format, in real time, throughout the simulator exercise. This information, together with the final report will be provide the detail required to enable the result of an assessment to be easily verified. Consequently, whenever the assessment is carried out, candidates can be sure that they are being measured against a consistent standard, and shipping companies can be more confident that those who meet this standard are more likely to be competent operators onboard their vessels rather than those who pass simple question type assessments or simulator based assessments which still impose a higher level of subjectivity in the final result.

Testing so far has been undertaken using with a model of an LNG (Membrane) Carrier, but by defining the appropriate knowledge and assessment database this methodology can clearly be applied to other ship types. As simulator assessment becomes more common, such as that required by Intertanko's TOTS scheme, use of this type of automated assessment tool will become more essential.

Acknowledgements

I would like to gratefully acknowledge the assistance of Ray Gillett, Director of Operations – UK, MPRI, and Tom Addis, Emeritus Professor, University of Portsmouth, UK., for providing the automated assessment tool for testing and their assistance in writing this paper.

Bibliography

Addis, T. R. and R. B. Gillett (2007). The Principles of Auto-Assessment**:** 1**-**26.

Dave, R. H. (1975). Developing and Writing Behavioural Objectives (Armstrong, R. J. ed). Educational Innovators Press, Tuscon, Arizona

Smith, K. and P. A. Hancock (1995). "Situation awareness is adaptive, externally directed consciousness." Human Factors **37**(1): 137-148.

Wolf, A. (1989). Can Competence and Knowledge Mix? Competency Based Education and Training. J. W. Burke, Routlidge.