Simulation Assessment Technology to Improve Objectivity, Evaluation Granularity and Opportunities for Data Analysis

Murray Goldberg

CEO, Marine Learning Systems

Vancouver, British Columbia

Murray@MarineLS.com

Arvinder Aujla

Marine Science Instructor

Esquimalt, British Columbia

Arvinder.Aujla@Forces.GC.ca

ABSTRACT

Advances in technology have enabled learning organizations to significantly improve knowledge acquisition and assessment. Many tools can be employed to test knowledge and can use the resulting data to pinpoint gaps in knowledge, as well as strengths and weaknesses in our training programs. For skills acquisition, advancement in simulation, serious games, augmented reality and virtual reality help to engage learners and optimize the learning experience. However, one significant gap remains: how to assess the performance of observable skills in a way that eliminates bias and delivers actionable metrics.

Assessing skill proficiency for individuals and teams is critically important, yet tools to support objective and consistent skill assessment are difficult to find. Skills assessment today is typically paper-based and relies on the viewpoint of expert assessors, limiting our ability to extract metrics to drive improvement. A wealth of information is lost and there is an incomplete view of learner abilities and skill proficiency.

This paper will provide an overview of new approaches to simulation assessment being used by the Royal Canadian Navy and CSMART, the world's largest civilian simulation training centre for mariners. These simulation centers have employed new technology that helps resolve the inherent challenges which exist in traditional forms of skill assessment. They are also developing insights-based models which ensure transparency and continuous improvement. The developments discussed here are applicable not only in the context of navy training, but also in operational assessment in the navy, the other branches of the armed forces, and more broadly in any context where skill proficiency and protocol adherence is important.

ABOUT THE AUTHORS

Murray Goldberg was a tenured faculty member conducting research on the effectiveness of web-based learning in the department of Computer Science at the University of British Columbia. Based on his research, in 1997 Murray founded WebCT, a Learning Management System that grew to serve 14 million students in 80 countries at 4,000 universities and colleges. In 2007, Murray became involved in maritime learning technologies through his work to help bring blended learning practices to British Columbia Ferry Services Inc. This led to the founding of Marine Learning Systems (MLS) where Murray serves as founder and CEO. MLS is the provider of enterprise learning technologies to cruise lines, ferries, coast guards, training centers and other segments of the maritime industry.

Arvinder Aujla served in the Royal Canadian Navy as a Naval Warfare Officer. After retiring in 2009 he accepted a position as a Marine Science Instructor at VENTURE. Arvinder seeks to leverage technology where it makes the training for Junior Naval Officers better.

Simulation Assessment Technology to Improve Objectivity, Evaluation Granularity and Opportunities for Data Analysis

Murray Goldberg
CEO, Marine Learning Systems
Vancouver, British Columbia
Murray@MarineLS.com

Arvinder Aujla
Marine Science Instructor
Esquimalt, British Columbia
Arvinder.Aujla@Forces.GC.ca

EDUCATIONAL TECHNOLOGIES FOR SKILL ASSESSMENT

Safe and efficient performance in a maritime setting requires knowledge and skill. Knowledge can be defined as the foundational or theoretical understanding of a subject and can be learned and transferred from trainer to learner. Skills are based on knowledge and are measured by our proficiency in applying knowledge to the subject at hand. Skills, like knowledge, can be learned and transferred from trainer to learner. It therefore follows that improving our ability to teach and assess knowledge and skills is central to our ability to perform correctly, safely and efficiently. Improving these training outcomes has been one of the central goals of educational technologies for the last three decades.

The best evidence of the ability for learning technologies to improve training outcomes is arguably a meta-analysis conducted by the United States Department of Education in 2010 (U.S. Department of Education Office of Planning, Evaluation, and Policy Development Policy and Program Studies Service, 2010). This analysis consisted of a systematic search of over 1,000 studies of online learning. There were several key findings. First, students learning online performed modestly better on average than students learning the same material in traditional face-to-face settings. Second, in blended environments where students had access to both online and face-to-face instructional elements, students performed better still. And notably, this effect was discipline agnostic. That is, the effect held regardless of what was being taught.

Thus, by the end of 2010, there was clear evidence that educational technologies not only improved access to training and assessment, but also supported improved outcomes. Not surprisingly, the deployment of these technologies continued to rapidly grow as a result. Now the use of educational technologies is measured in market size by the billions of dollars. It is projected to increase to over 400 billion dollars by 2025 (Rivero, 2020).

The foregoing growth primarily refers to what one might think of as traditional educational technologies. That is, the now-ubiquitous Learning Management System (LMS) that supports and tracks both the learning and assessment of knowledge.

But what about the learning and assessment of skills? It is well understood that the effective performance of skills, especially in unpredictable or varying situations, is underpinned by a solid understanding of the knowledge at the foundation of that skill. However, while skill-associated knowledge is necessary for optimal skill performance, it is not sufficient. As such, a robust set of technology-based tools supporting the learning and practice of skills also emerged in the form of visualization and simulation technologies. Simple computer-based simulation technologies arrived as early as the late 1940's with the development of general-purpose computing (Goldsman, Nance and Wilson, 2009). The use of simulation expanded starting in the 1970s as enhanced analytical and modelling tools became available. Now the use of simulation technologies is very common in many high-stakes environments such as medical surgery, aerospace, and of course maritime navigation, engine operation, bridge resource management, and elsewhere. As augmented and virtual reality technologies continue to advance, simulation technologies benefit by becoming less expensive, more highly available, and more realistic (Mallam, Nazir and Renganayagalu, 2019).

Simulation offers the advantages of being able to place learners in environments which are highly controlled and are unlikely to commonly occur in a natural setting. This has obvious advantages for improving outcomes when personnel are faced with a situation that might occur once in a lifetime but have been experienced dozens of times in simulation. In terms of quantifying the effectiveness of simulator training, there have been a wide variety of studies including (Sangeetha and Gomathy, 2019). As is the case for LMSs to support knowledge learning and assessment, bridge and engine room simulators have become commonplace in maritime skill training (Mallam, Nazir and Renganayagalu, 2019).

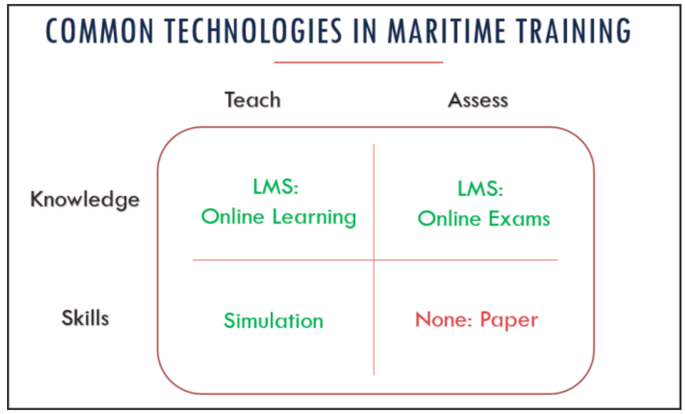


Figure 1: Learning and Assessment Technologies

Together, the foregoing speaks to the progress and impact of technology on knowledge learning, knowledge assessment, and skill learning. However, skill assessment has been left behind and in contrast to the other three quadrants of the training matrix, skill assessment typically remains paper-based (see Figure 1). Given how technology has improved training in the other three quadrants, it is natural to question and investigate whether technology can be leveraged to improve upon the existing state of simulation assessment. The remainder of this paper describes such a project being undertaken at the Canadian Naval Officers Training Center (NOTC) in British Columbia, Canada.

NAVIGATION AND BRIDGE SIMULATOR ASSESSMENT PRACTICES

Introduction

The Canadian Naval Fleet School Pacific houses VENTURE Division. VENTURE is located at Work Point in Esquimalt, B.C. and is dedicated to the professional education and development of Canada's future naval leaders. In addition to Naval Warfare Office (NWO) core occupation training, VENTURE also conducts Fleet Navigating Officer, Arctic Operations (Navigation Phase), Command Development and Ship Handling courses. VENTURE also supports Canadian Fleet Pacific (CANFLTPAC) through refresher training and command mentorship.

Using a wide range of modern facilities, VENTURE teaches a balance of the theoretical and practical attributes and skills required by naval officers. It was in the late 1990s that a reduction of sea-going platforms saw the impetus for the Canadian Navy to undertake a trial where the possibility of virtual reality simulation (VRS) was explored. Over a period of time and many trials it was proved that simulation could be an effective training tool in a modernized curriculum. Today, Naval officer training is conducted within the Vice-Admiral A.L. Collier Building, which houses electronic classrooms for academic instruction, the world-class Navigation and Bridge Simulator (NABS), the Naval Part Task Trainer (NPTT) and two multimedia theatres.

The NABS facility is a state-of-the-art visual and radar simulator. The NABS at Naval Fleet School Pacific NFS(P) has one 360° trainer and six 320° trainers. Each trainer is able to provide accurate, challenging, and realistic scenarios in support of NWO training. Areas of training include Officer of the Watch (OOW) development, execution of navigation, the conduct of shiphandling (such as replenishment at sea or RAS), helicopter operations, alongside and departure manoeuvring, towing at sea and Fleet Manoeuvring. The NABS architecture can be configured to emulate any major platform in the Canadian fleet. Each platform incorporates an accurate physical and hydrodynamic model so as to respond with great fidelity. The NABS operators (all ex-naval officers with at least one Command tour at sea) in the control room can manipulate the system to introduce faults such as steering gear breakdowns, engineering emergencies, or faults in the electronic chart display, radar and gyro suites. They are able to introduce and manipulate any and all environmental conditions such as wind, fog, and current. The greatest strength of the NABS is in its ability to provide the learner with an exceedingly high degree of interactivity based on the learner's real-time input.

The NWO program utilizes simulation training as part of the training continuum which is broken into three phases: fundamental (NWO II), core (NWO III) and advanced (NWO IV). The entire training cycle has a duration of 205 training days, and each cycle consists of an academic or classroom phase, a simulation phase and finally an at-sea phase.

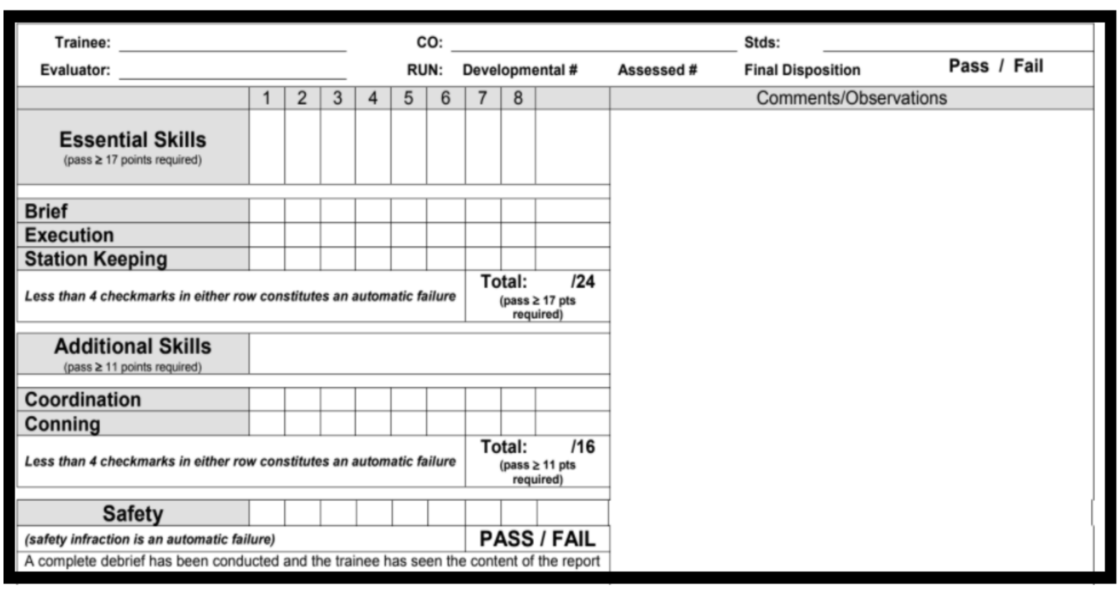
The emphasis of this paper is the simulation phase of the core (NWO III) training. The simulation phase commences with a series of scripted developmental runs of equal duration where the student is exposed to the prescribed learning outcomes in accordance with a qualified standard plan. The student is coached by both a Mentor (a retired Naval Officer who has held Command at Sea of a major combatant) and a Course Training Officer (CTO - a serving line NWO usually of the rank of Lt(N) who is highly proficient and conversant with current standard operating procedures (SOPs) employed by the Fleet. At the end of each developmental run the student receives an assessment sheet where their performance is measured against a rubric. The student is shown their areas of strength and more importantly where improvement is required to meet the standard. After the period of development consisting of a set number of runs, the program shifts to an assessed phase where the Mentor and Course Training Officer assess the performance with a pass-fail adjudication.

The NWO III phases consist of assessments in the areas of OOW, OOW Manoeuvring (Fleet Maneuvering) and navigation.

NABS Assessment Methodology and Outcomes

The current method to capture student performance is through hand-written 'clipboard' assessment sheets. The sheets are designed with various fill-in-the-box data entries such as surname, type of run, run number, ephemeral data around weather conditions (day run, night run, visibility, wind, current etc.). The main section contains key performance indicators (KPIs) which are measured against a rubric for consistency (see Figures 2 and 3). There is also the ability for the CTO to provide freeform observation to augment the assessment. The final section delineates the disposition of the candidate.

A typical course has a cohort of 20 students. With four NABS bridges running each with five scheduled daily runs, it allows each student to have one run per day generating 20 assessments. Thus, in one week there are approximately 100 assessments, one for each run conducted.



Trainee: ______ CO: ______ Stds: ______

Evaluator: ______ RUN: Developmental # Assessed # Final Disposition Pass / Fail

	1	2	3	4	5	6	7	8		Comments/Observations
Essential Skills (pass ≥ 17 points required)										
Brief										
Execution										
Station Keeping										
Less than 4 checkmarks in either row constitutes an automatic failure							Total: /24 (pass ≥ 17 pts required)			
Additional Skills (pass ≥ 11 points required)										
Coordination										
Conning										
Less than 4 checkmarks in either row constitutes an automatic failure							Total: /16 (pass ≥ 11 pts required)			
Safety										
(safety infraction is an automatic failure)							PASS / FAIL			
A complete debrief has been conducted and the trainee has seen the content of the report										

Figure 2: Assessment Sheet

Brief:

The brief to the CO should be:

- Logical; including a feasible solution and factually accurate
- Delivered in the prescribed format
- Contain relevant safety information ***Key ranges and bearings may be included in the brief but are not required for a satisfactory brief.***

Execution:

Solutions should be applied without delay after receipt of the execute signal. To satisfactorily execute a manoeuvre, the trainee must complete the following:

- Come to the solution course and speed
- Check safety. Specifically that consorts included in safety considerations are drawing the anticipated direction
- Use rel vel where appropriate to update the solution
- Monitor key bearings and ranges where required and apply them correctly

Station Keeping:

Maintained a position within 50 yards of the assigned station. Demonstrated an understanding of Circle principle by:

- correctly assessed station and,
- applying an appropriate correction when required

Coordination:

Managed the flow of information from the various members of the team and provided timely directions to the team. This includes:

- Managing the NAV COMM and tasking the rel vel officer appropriately
- Giving direction to the range officer for key ranges and required range reporting
- Prioritizing the various inputs
- Expeditiously working out the details for the next manoeuvre without losing focus on the present manoeuvre

Conning:

- Checked the ship's quarter in the appropriate direction before applying helm
- Applied the correct helm and steadied the ship's head within 5° of the intended course
- Student ***shall*** check the rudder angle indicator after giving any conning order. If it is not checked, this qualifies as a safety infraction and a failed run.
- Used approved conning orders as detailed in Ships Standing Orders

Safety :

A close quarters situation is deemed to exist when two ships approach each other within 200 yards. If the actions of the trainee create a close quarters situation, during which two or more consorts approach within 200 yards of one another, then the trainee will fail for safety. If however, a student is involved in a close quarters situation which they did not cause, they may not fail provided the following conditions are met:

- The trainee recognises the situation developing and, once the consort comes within 300 yards, briefs the CO including range and bearing trend (drawing left/right closing)
- The trainee makes a recommendation to the CO to avoid collision. Even if this recommendation is not correct or may not resolve the issue. The CO will assist the trainee if the trainee's recommendation is deemed unsafe. A student will not fail based on the validity of this recommendation alone.

Even if two or more ships do not approach within 200 yards, a safety infraction may be deemed to exist if the actions of the trainee place their own ship or a consort in extremis or if the rudder angle indicator is not checked after every conning order as detailed above in "Conning".
Discretion rests with the CTO to determine whether a trainee's actions or inactions resulted in a safety infraction and whether the circumstances warrant clemency. Finally, the CO must be alerted during the initial brief as part of the safety considerations if coming within 200 yards of a consort is expected as part of the solution (e.g. Station J) or if a close quarter situation may result from a consort being out of station or some other happenstance involving a consort (wrong or inappropriate solution, wrong conning order).

Figure 3: Assessment Rubric

Limitations of Current Assessment Practices

There are inherent limits to the value that can be derived by these assessment practices.

First, assessment forms, as illustrated by the example provided above, tend to be neither highly specific nor comprehensive in their rubric or description of scoring requirements. Instead, there is a significant degree of expertise assumed on the part of the assessor in order to simultaneously record the assessment observations and interpret those observations in the process of deriving a scoring outcome. On the one hand, this could be an asset in that an assessor has latitude to contribute value to the assessment by leveraging his or her knowledge and experience and using those to inform the assessment outcome. On the other hand, this lack of specificity and comprehensiveness creates the requirement for significant expertise on the part of the assessor in order to produce valid assessment results (where the assessment actually assesses what it is intended to assess). However, expert assessors are often a scarce resource limiting opportunities for formative assessment. Given that formative assessment in various disciplines has been shown to materially improve outcomes (especially, and interestingly, computer-based formative assessment systems) (Kingston and Nash, 2011), reducing the reliance on expert assessors may result in more opportunities for formative assessment, and thus in improved outcomes.

Second, this reliance on expert assessors in place of highly specific and comprehensive assessment instruments creates the requirement, as noted above, for the assessor to use his or her expertise to inform the assessment results. Given the natural variability from one assessor to another regardless of their level of expertise, subjectivity is introduced creating variability in assessment outcomes. Nuances noticed by one assessor may be missed by another. Assessors may have differing views on acceptable techniques or what constitutes satisfactory vs. exemplary performance. This impacts a core requirement of quality assessments called reliability - the ability of an assessment to consistently measure student performance.

Third, because current assessment methodologies are typically paper based, the assessor is essentially acting as the "computer" given responsibility for distilling observations into an assessment outcome. This precludes the utilization of more complex and potentially beneficial algorithms to derive an assessment outcome from the given observations. Not only is this assessor-based "computation" limited in the complexity it can accommodate compared to technology-enhanced computations, but those computations can only utilize the data available on the assessment form or in the head of the assessor. This introduces subjectivity and imposes limits in the depth of the derived assessment outcomes, as each assessor is relying on a different, inherently limited set of lived experiences as their basis for comparison against the performance being assessed.

Finally, also due to the paper-based nature of current assessment practices, there is a limit on the value of the data collected beyond an immediate debrief. If, instead, historical assessment outcomes were logged into a database, significant additional opportunities would emerge to analyse trends and gaps both for individual trainees as well as for the workforce as a whole. The more granular the historical assessment data available, the deeper the insights that might be derived.

These realities are well understood by program leaders. A survey of NABS CTOs and mentors revealed that nearly all felt that simulation assessment required a very high degree of expertise on behalf of the assessor. Yet roughly 25% could not agree with the statement that paper-based simulation assessment was highly objective, nor could they agree that the results were highly consistent. Fewer than 50% agreed that current paper-based practices allowed them to identify trends and gaps over many assessments. Fewer than 30% agreed that current assessment practices allowed them to concretely say whether training practices were more or less effective than those in the past and why. Fewer than 40% agreed that current assessment practices enabled them to easily rank trainees, past and present, for a given role. And finally, only a little over 40% could agree that current assessment practices are impartial and data-driven, and even fewer felt that trainees saw them as such.

Taken together, all the above indicate that there may be a significant opportunity to improve the value of our simulation assessments by reconsidering the current practice of paper-based assessments. By employing purpose-build simulation assessment technologies, there is the potential to increase opportunities for formative assessment, increase the objectivity and consistency of assessment, increase the sophistication of assessment algorithms, and derive descriptive, diagnostic, predictive and prescriptive analytics from our assessment data addressing both individual learners and our workforce as a whole.

TOWARDS SKILL ASSESSMENT TECHNOLOGIES

Introduction

The aforementioned shortcomings of current, paper-based assessment practices were noted in 2017 by CSMART, a large maritime simulation training center in Almere, The Netherlands. CSMART is owned by the Carnival Corporation & PLC, the largest cruise line company in the world, and the operator of the Carnival, Holland America, Costa, Aida, Princess, Seabourn and other cruise lines. CSMART is an advanced training center whose mission includes not only course delivery, but research, development, and quality assurance. They have pioneered simulation training advances including a role-based approach to bridge management and engineering officers. Their simulation facilities include four full-mission bridges, eight part-task bridges, four full-mission engine room simulators, and twelve virtual machinery space stations.

Having continuous improvement as a core value, in 2017 CSMART identified the opportunity to improve the quality and value of simulation assessments. Working with their learning technology partner Marine Learning Systems (the company) and external human performance experts, the group defined and then constructed a prototype skill assessment technology to address the above-mentioned shortcomings of current practice. Although the prototype was designed only as a proof of concept to validate the opportunities for improving assessment, it was very well received and soon replaced paper for nearly all assessments conducted at CSMART. Ultimately, a tablet-based application was developed based on the lessons learned from the use of the prototype. It is this skill assessment application, called SkillGrader (the application), whose use at the Canadian Navy Navigation and Bridge Simulation (NABS) Center is described in the upcoming sections of this paper.

SkillGrader Application, Context and Use Case

The application is a tablet-based application for assessment and debrief, which is linked to a back-end server for data storage and analysis (see Figure 4). The assessment scenario is unchanged when using the application; an assessor observes one or more trainees performing a simulation exercise. However, instead of a paper-based form for recording the assessment, the assessor fills out an electronic “form” on the application. As in the case of paper-based assessment, there is a separate electronic form for each assessment scenario. But unlike the paper-based case, the form elements on the application consist only of binary yes/no observations to indicate whether a form element, called a “Performance Indicator” was observed being performed by the trainee. Whereas paper-based forms might consist of several Likert-based scores supported by a rubric to record various activity outcomes, performance indicators on the application are intended to embody the rubric and derive the score electronically.

For example, on a paper-based form, the rubric might indicate that the trainee should be awarded 3 out of a possible 5 by the assessor if they perform three specific actions. The application, instead, lists performance indicators for the potential actions performed by the trainee, and if they were observed, calculate the score of 3 automatically. This example extends in scope and complexity to various competencies being assessed by the form, and to the form in its entirety. Performance indicators are tagged according to the competencies they measure (e.g., navigation, leadership, etc.), their level of importance to that competency (critical, important, mastery, etc.), and their weight or degree to which they contribute to measuring performance (e.g., a weight of 5 out of a possible 10). From these tags, the observations of the assessor, and the configurable threshold-based scoring algorithm, scores are generated for each trainee, both for their performance as a whole, and for each competency measured by the exercise.

Collectively, the above is intended to improve assessment objectivity, consistency and debrief detail. Assessors are more likely to agree on whether a performance indicator was observed than whether the performance merits a 3 out of 5 vs. a 4 out of 5. Likewise, the complex calculation of performance both overall and for each competency measured can be more fine-grained yielding more detailed metrics than possible by hand. And finally, these calculations are performed identically, in every case.

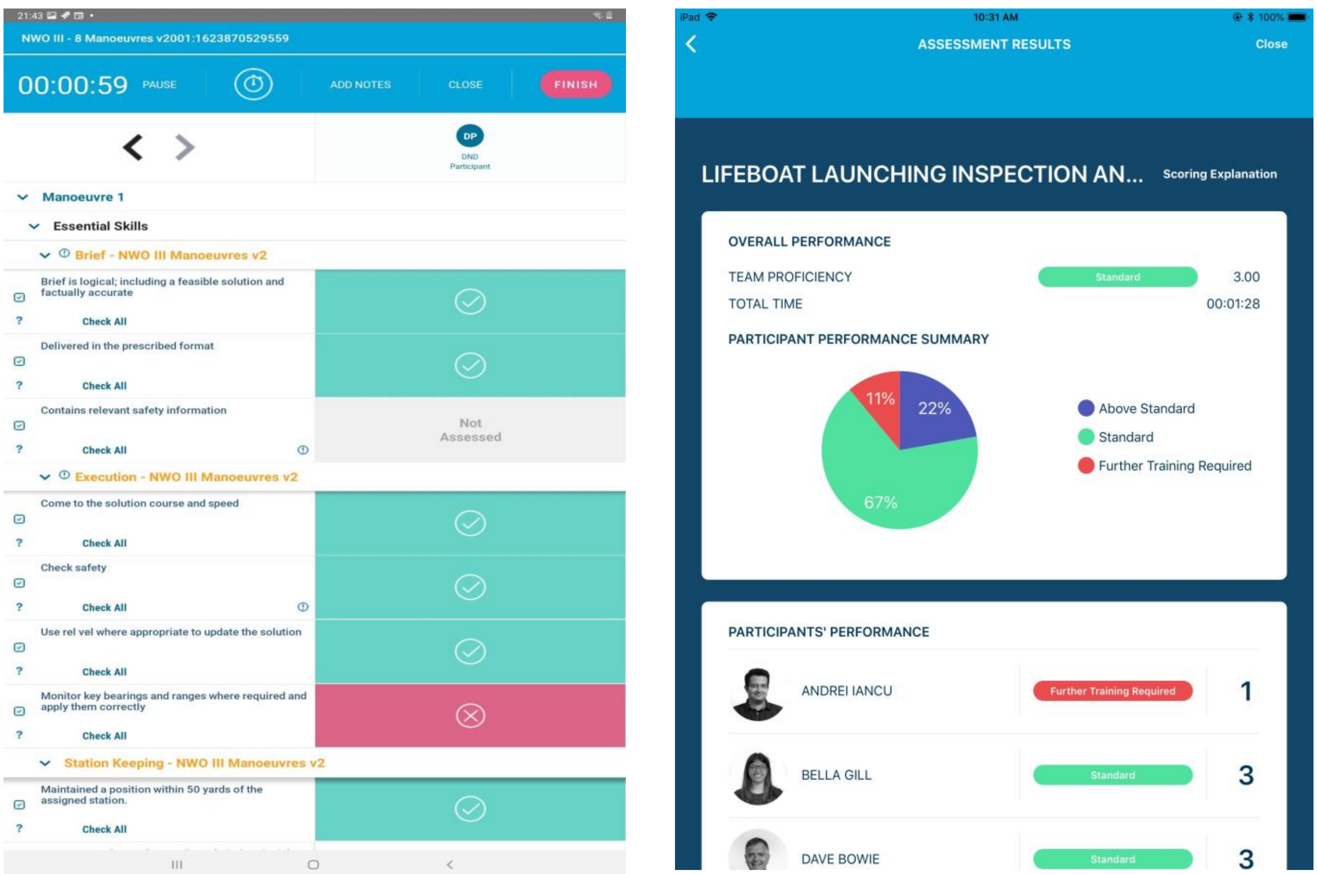


Figure 4: Sample Application Assessment Form and Debrief Report

Immediately upon conclusion of the assessment, the application displays a performance outcome screen to facilitate debrief. This screen displays the overall score, as well as a score for each competency measured. It is being extended to add additional visualizations such as a spider chart, etc. to further support a comprehensive debrief.

Learning Analytics Opportunities

Also, upon conclusion of the assessment, the application transfers all of the raw performance indicator data and the calculated scores to the back-end server in support of further analysis. The assembly of this assessment data over time creates the opportunity to gain insights and identify trends more easily than is the case in paper-based analysis. For example, fine-grained assessment data enables a deeper analysis of the performance of the individual learner inside a course, for the duration of the time at the school and perhaps even over a career. Overall strengths and weaknesses from course to course can be detected allowing for the development of an individual learning plan to address shortfalls in particular domains such as navigation, bridge skills, or interpretation of rules of road.

At the next level these analyses enable intra-cohort analysis where the performance of the cohort is analysed as a whole against other cohorts. This creates the opportunity to detect course instance issues such as CTO differences, mentor contributions and influences, training time-of-day effects, or contextual issues with varying training flow and external workload.

At a broader level inter-cohort analysis can present an opportunity to reveal program-wide insights such as common gaps, common strengths, timing of skill fade and others.

These analyses which take as input the data produced by the application and other learning records are an example of learning analytics, a data science approach to analysing large volumes of data to determine trends, correlations and other insights (see Figure 5). Learning analytics can be retrospective or predictive. That is, we can look at past data to tell us what happened and why it happened. But perhaps more interestingly they can help us determine what is likely to happen in the future based on past events. Gartner has broken this down into four kinds of analytics. "Descriptive analytics" tell us what happened. "Diagnostic analytics" tell us why something happened. "Predictive analytics" inform us of what is likely to happen. And finally, "prescriptive analytics" help us understand how we can make a desired goal happen in the future. The nature of the application data is especially useful from an analytics perspective. Because it can generate objective "real world" skill performance data, this enables the discovery of correlations between the inputs (trainee background, courses and grades, instructors and mentors, sea time, etc.) and the outputs (skill assessment data). This can yield information such as how effective is each course in determining a particular aspect of performance? What is the relative effectiveness of each instructor? What is the optimal retraining interval? Who, amongst our learners, are the most likely to emerge as leaders, or most at risk of failure? And in general, how can we improve our training and recruitment practices to optimize workforce performance.

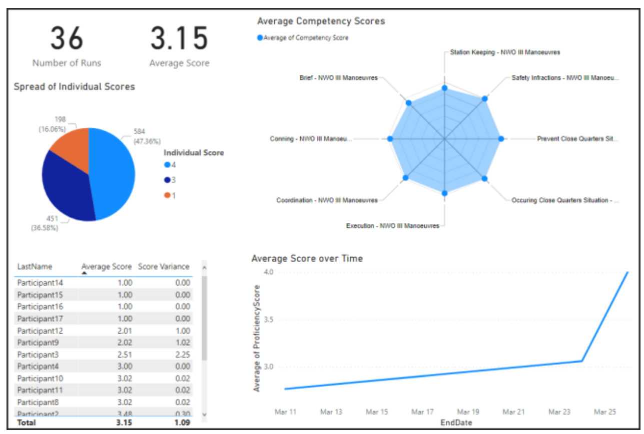


Figure 5: Learning Analytics Example

EARLY EXPERIENCE WITH THE APPLICATION IN NABS ASSESSMENTS

The leaders of the NABS Metric Project at the NFS(P) were interested in determining how analytics could enhance the current program of NWO simulation training. It was decided to take a step-by-step approach.

The first was to take the current paper assessments as a starting point and digitize the process of collecting data. The digitization of data immediately would create a repository of data. Once established, data science approaches could be considered.

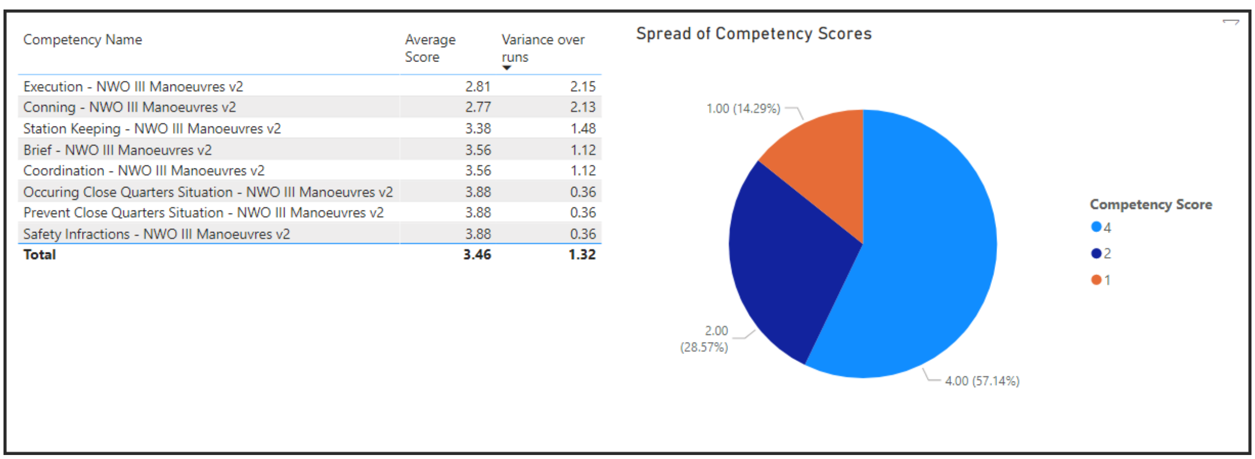
One initial challenge was to ensure that the rubrics associated with the assessments would lend themselves to digitization. The current assessment form consists of a combination of assessment ephemeral data (name of student, course details, the run details, developed or assessed run), along with key performance indicators (KPIs) graded against a rubric standard, leading then to a final pass or fail disposition.

Digitization of forms in five different courses was undertaken. The courses were the following: the Naval Warfare Officer Phase III, Naval Warfare Officer Phase IV, Fleet Navigation Officer, and the Command Development Course. Forms assessing navigation, bridgemanship and fleet manoeuvres were developed for the application. The process led to the realization that while the existing paper-based forms capably determine whether the student met the standard, they are generally unable to identify a gradient of performance for those that are above the standard.

Another issue raised by the introduction of binary performance indicators in the application over rubric-informed grading by assessors on paper, was that of subjective vs. objective assessment. While in general objective measures of assessment are favoured for their fairness, consistency and reliability, it can be argued that some measures of performance cannot be measured objectively but instead require the intuition of experienced assessors to properly evaluate. The electronic approach of the application does not preclude the use of subjective measures of assessment, but it does stimulate the discussion of objectivity vs. subjectivity in general, and thus forces a conscious choice to be made between the two when designing assessment forms for the application. This was viewed as a healthy, overdue discussion.

The introduction of the application has not been exempt from the normal considerations of change management in a secure environment. To address these issues, first a robust communication plan was developed. The first framework of the plan centered on the implementation of the technology. It required a certification process that would allow a 'closed loop' system to be established on the footprint of a defence establishment. Tablets capable of having direct communications through a secure Wi-Fi channel to the backend repository were configured. The second framework addressed change management for the human element. This was created to ensure the CTOs, Mentors and students understood the basic project details, how it would bring value to the current programme, how it would be introduced, and finally what the role of the CTO and the Mentor would be with regard to moving from a 'paper-pen and clipboard' to a 'digitized-tablet' format. The challenge here was and is the momentum created by the strong organizational culture ensconced over 23 years, and the effort and trust required to consider and then actually change the manner that assessments are carried out.

Although very early, application analytics have already revealed interesting results (see Figures 6, 7 and 8). For example, the data has revealed which areas of bridgemanship our trainees are weaker in and stronger in. Although some of this may have been felt intuitively by our assessors, access to hard data allows us to consider concrete changes to training such as the creation of part-task scenario training to address common areas of weakness.



Competency Name	Average Score	Variance over runs
Execution - NWO III Manoeuvres v2	2.81	2.15
Conning - NWO III Manoeuvres v2	2.77	2.13
Station Keeping - NWO III Manoeuvres v2	3.38	1.48
Brief - NWO III Manoeuvres v2	3.56	1.12
Coordination - NWO III Manoeuvres v2	3.56	1.12
Occuring Close Quarters Situation - NWO III Manoeuvres v2	3.88	0.36
Prevent Close Quarters Situation - NWO III Manoeuvres v2	3.88	0.36
Safety Infractions - NWO III Manoeuvres v2	3.88	0.36
Total	**3.46**	**1.32**

Figure 6: Competency Performance and Variance

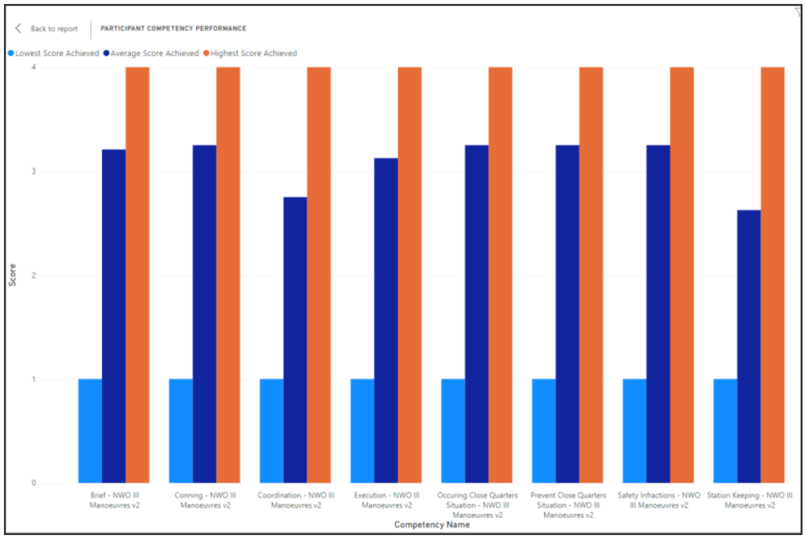


Figure 7: Competency Min, Avg, Max Performance

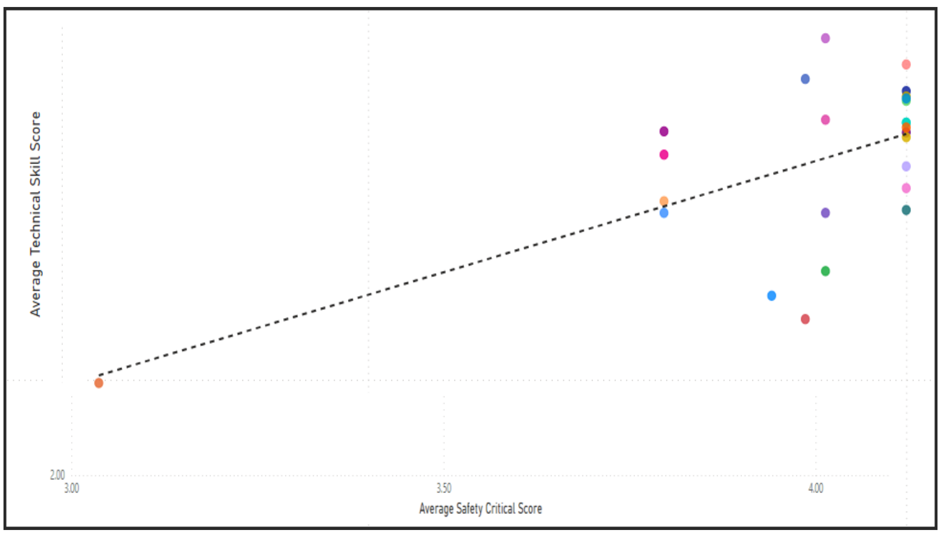


Figure 8: Technical vs Safety Score Plot

Another very interesting result is variance in competency as revealed by the analytics. That is, although it is expected that some competencies are generally performed better or worse than others by most students, what was not previously clear was that there is greater variation in the performance of some competencies than others. Understanding the degree of variance in each competency allows us to choose which competencies to correlate with inputs (such as instructors, previous experience, etc.) in order to understand what inputs produce the best outputs. Variance also presents opportunities for streaming trainees into different operational paths, and to understand how the performance of groups of competencies vary together, or inversely. All of this allows us to measure the success of our training in terms of the performance it produces, to deeply understand that which was only superficially understood before, to explore opportunities to identify and close gaps for individuals and for the program as a whole, and to use precision data to identify, support and stream trainees throughout their career in the armed forces.

Looking Forward Within the Navy, More Broadly in the Armed Forces, and Elsewhere

The early use of the application in the NABS program raised several opportunities to improve the application for assessments of this type. For example, it became clear that assessments are dynamic in terms of how many navigation runs might be experienced, with the number not necessarily being known at the outset of the assessment. Therefore, there is work underway to vary the number of assessment phases in situ, during an assessment. Likewise, the ability to take notes and add labels as additional descriptive data during an assessment became clear. There were also some simple user interface improvements suggested. In all, the work to introduce the application to the NABS assessment processes has been highly valuable in improving the utility and value of the application.

From the NABS program perspective, the current phase is only an initial familiarization, establishment and optimization phase. The hope and expectation are that this will result in a tried and tested pathway to the new model for assessing students. Once done, this potentially opens the door to the use of this technology in assessment outside the NWO courses.

Although this work addresses a skill assessment during training, the discussion is equally applicable in operational contexts, in other branches of the armed forces, and elsewhere. Generally, the application can be used in any domain to assess individual training (IT) events and team training (TT) events. It is also applicable in force generation (FG) events and force employment (FE) events. Skill assessment forms for the application could be adapted to perform IT and TT assessments in damage control operations, sea going flight operations, warfare serials such as gunnery, and anti-submarine operations. Very similar assessment requirements are seen in the army and air force, and assessment forms could be created to assess training and operations in those domains as well. It is our hope that we, or others, will expand this work into all these arenas as applicable.

REFERENCES

Goldsman, D., & Nance, R., & Wilson, J. (2009). A Brief History of Simulation. *Proceedings of the 2009 Winter Simulation Conference*, 310-313.

Kingston, N., & Nash, B. (2011). Formative Assessment: A Meta-Analysis and a Call for Research. *Educational Measurement: Issues and Practice, Volume 30, Issue 4,* 28-37.

Mallam, S., & Nazir, S., & Renganayagalu, S. (2019). Rethinking Maritime Education, Training, and Operations in the Digital Era: Applications for Emerging Immersive Technologiesation. *Journal of Marine Science and Engineering*, 7*(12)*.

Rivero, V. (2020). *Growth of edtech through 2025*. EdTech Digest. https://www.edtechdigest.com/2020/11/13/growth-of-edtech-through-2025/].

Sangeetha, V., & Gomathy, D. (2019). Effectiveness of Seafarers Training Using Maritime Simulators. J*ournal of Management*, 14-20.

U.S. Department of Education Office of Planning, Evaluation, and Policy Development Policy and Program Studies Service (2010). *Evaluation of evidence-based practices in online learning*. U.S. Department of Education.