Submitted for January-March 2000 issue of
Proceedings of the Marine Safety Council on the implementation of the
International Convention on Standards of Training, Certification and Watchkeeping (STCW)

How Do You Assess Mariner Proficiency?

By Dr. Myriam Witkin Smith and Dr. Marvin C. McCallum

The 1995 amendments to the International Maritime Organization's (IMO) Seafarer's Training, Certification and Watchkeeping (STCW) Code demands many changes in the way the maritime industry assesses the proficiency of its mariners. The industry does not have much experience with assessment by practical demonstration before an expert assessor and there are many questions to be asked during the adjustment to these new requirements. In 1997 the U.S. Coast Guard (USCG) Research and Development Center (R&DC) and the USCG National Maritime Center (NMC), together, selected some key questions presented by the STCW Code:

- What does a valid and reliable assessment of mariner proficiency look like?
- Can industry instructors and examiners prepare such assessments?
- What are the special requirements of using simulators for assessments?
- What are the special requirements of conducting assessments in the operational shipboard environment?

What does a valid and reliable assessment of mariner proficiency look like?

To answer the first question, we brought together a team that, several years before, had successfully developed an objective assessment of Rules of the Road knowledge and skills. Our team of human performance experts from the R&DC and Battelle Seattle Research Center (BSRC) and marine educators from the U.S. Merchant Marine Academy (USMMA) carefully examined the requirements of the STCW Code for assessment. We supplemented those requirements by consulting the Instructional Systems Development (ISD) methods used by the military and by industry and, most especially, by examining the best practices of the maritime industry. What we learned was that an assessment should be a *reliable* and *valid* snapshot of the performance that can be expected from a mariner. A "reliable" assessment is one whose consistency can be trusted: the same performance will receive the same assessment every time. A "valid" assessment is a sample of performance that includes all the critical components of the function that will be expected from the mariner on board the ship. From what we learned, we designed a step-by-step method for developing reliable and valid assessments.

The first step of our method is to specify the *assessment objectives*, a listing of the knowledge, skills, abilities, and proficiencies that represent the critical requirements for the competence of interest. Preparation of this list provides an opportunity for review by industry experts and provides a foundation for further development of a valid assessment procedure. With the objectives specified, it is possible to determine the *methods* for the assessment and to specify the *conditions* under which the mariner will demonstrate

proficiency. The least intuitive step in the method requires development of the performance measures and standards, which are critical to the reliability of the assessment. *Performance measures* are observable behaviors or the observable consequences of behaviors; *performance standards* are acceptable or target levels to be achieved by the observable behavior or consequence. For STCW implementation, the standard is intended to define the minimum acceptable level of performance. An example of an objective might be to assess a mariner's ability to steer by gyrocompass. The corresponding observable measure might be the accuracy with which the mariner is able to maintain the ordered heading. The standard might be to maintain the ordered heading to an accuracy of plus or minus three degrees. To continue this example, if the mariner can achieve that accuracy of steering, he/she is to be considered minimally proficient in steering by gyrocompass.

In order to test and refine our method, we did a case study. We selected an assessment emphasized in the STCW Code, the competence of an officer in charge of a navigational watch in the use of the Automatic Radar Plotting Aid (ARPA) to maintain the safety of navigation. We depended heavily on the assessment objectives defined in the Code and concentrated our attention on the later steps of the development process. We were fortunate in having faculty from the USMMA as members of our team. They provided their expertise in marine training and assessment, in simulator application, in ARPA, and, as a bonus, access to their fully capable ARPA laboratory based on stimulated real equipment. The general method our team developed is described in several project reports and in IMO Advisory Circular Number 853. The ARPA assessment developed in our case study is also available in a project report for industry review and for potential use.

Can industry instructors and examiners prepare such assessments?

We had demonstrated the feasibility of our method under fairly ideal conditions: a multi-disciplinary team, with dedicated time, with a fully capable simulator, and a competence thoroughly described in the STCW Code. Was our method useful to a broader segment of the industry? How much would it need to be refined? What kinds of materials would "qualified instructors," developing assessments, need to guide them? In the Spring of 1998, we hosted two workshops for a variety of instructors from marine academies, simulator facilities, large and small training schools, and large and small shipping companies. We were gratified by the readiness of so many people to participate in the project and by their generally positive response to the method. During the workshop, we presented the method, conducted a series of practical exercises, and obtained feedback from participants. We immediately learned that we had more work to do. Our workshop materials where not ready to support assessment development without help from human performance experts. We had included too much confusing terminology from the STCW Code and from various versions of ISD. And the problems of *developing* assessments and of *conducting* them were sufficiently different to deserve separate treatment.

For further refinement of the method for developing assessments, we were fortunate to be able to enlarge our team with faculty from two state academies. Faculty from Massachusetts Maritime Academy (MMA) provided new case studies on navigational

watch procedures for *Lookout* and *Helmsman*. While they developed their assessments and commented on our supporting materials, we commented on their assessments and developed our supporting materials. The language in our materials was simplified, giving priority to that of the STCW Code, and we prepared a *Developing Mariner Assessments* manual. For further refinement of the manual, faculty from California Maritime Academy (CMA) developed assessments for engineering watch procedures, *Preparing the Main Engine for Operation* and *Locating Generator Faults*. We found that industry instructors can, indeed, develop valid and reliable assessments. The manual, other materials, and the sample assessments that our team developed are available in project reports.

What are the special requirements of using simulators for assessments?

Our team had the advantage of the USMMA's highly capable ARPA simulator laboratory when we developed our ARPA assessment. However, at the present time, there is a broad range of simulators available to support any given mariner assessment. With the advances in personal computer (PC) processing capability, the advantages of more elaborate simulators are becoming less pronounced. One of our objectives was to develop and test a method to evaluate a simulator's capability to support mariner assessment. While doing that, we continued our ARPA case study. How well could two sample off-the-shelf PC-based simulators support the *mariner assessment objectives* that we had already developed? Our *simulator evaluation objectives* were based, first, on the mariner assessment objectives developed for ARPA operation. The simulators were evaluated on their capability to provide the prescribed exercise conditions needed for the mariner to realistically demonstrate the performance to be assessed. Then, we added additional evaluation objectives based on the STCW Code's standards governing the use of simulators. We designed an evaluation protocol, applied it to the two PC based simulators, and analyzed the results. We found that both simulators satisfied many, but not all, of our objectives and would have to be augmented by other forms of assessment. More importantly, we demonstrated the feasibility of a rigorous evaluation of simulator capability. Our general evaluation method and our protocol for evaluating ARPA simulators are available in a project report.

What are the special requirements of conducting assessments in the operational shipboard environment?

The most demanding setting for an assessment is on board a commercial vessel where it cannot be allowed to interfere with the safety and efficiency of operations. We were again fortunate in extending our team and in being able to examine assessment in this environment. SeaRiver Maritime, Incorporated (SRM) had sent a representative to our earlier workshop and in 1999 agreed to participate in the development of a "package" of documentation for the shipboard assessor and to actually conduct a series of trial assessments on board their ships. They began by reviewing the sample assessments for navigation and engineering watch competencies that had been developed by the Academies. During these preliminary preparations, the first issues identified were the amount of detail needed in the onboard assessment package and the amount of training

needed by the onboard assessor. A related issue was the degree to which the Academies' assessments needed to be adapted to the operating procedures of the particular company and ship operations.

In late 1999, extensive trials of the assessment approach were conducted onboard SRM tankers. Officers who augmented the ship's regular crew conducted the earlier trials while BSRC experts observed. As the assessments became more polished, the experienced officers introduced the procedures to regular ship officers for them to apply. Figures 1 and 2 are photographs of trial applications of the assessments onboard SRM vessels. As of this writing, only preliminary results of the trials are available. We found that the extent to which assessment procedures require modification for use on a particular ship depends on the general type of assessment. Assessments that address general competencies, such as navigation watch *Lookout* and *Helmsman* procedures, can be used on similar vessels with minimal modification, as long as the performance standards do not conflict with the ship's standing orders and normal operating procedures. However, assessments that address competencies involving vessel-specific equipment, as was the case with the *Prepare Main Engine for Operation* and *Test the Steering Gear* procedures, must be tailored to a vessel's equipment and operating procedures. In addition, we found that it was difficult for some of the regular ship officers to complete some of the assessments during the limited time period provided during these trials. This suggests the need to more completely integrate the assessment process into current operations and training, as well as to refine the assessment procedures to better match shipboard operational conditions and constraints. The results from conducting the trial assessments and the material developed will be available in later project reports.

Figures 1 and 2 about here

Project conclusions

Our project served as a laboratory for the USCG and the maritime industry to examine some of the key issues in understanding and implementing the recent requirements of the 1995 Amendments of the STCW Code. With USMMA, we discovered what was needed to develop assessments that were fully compliant with the STCW Code and that were reliable and valid by the best principles of ISD. With MMA and CMA, we demonstrated that qualified instructors could develop such assessments with appropriate guiding materials. We demonstrated an approach to evaluating simulators in their capability to support assessments. With SRM, we are in the process of determining what is needed for conducting assessments onboard commercial vessels.

Our project reports contain three types of "products." The first are our "laboratory" explorations of some of the critical components of the STCW requirements. We followed the guidance for assessment activities in that document, reported our experiences, and shared the lessons we learned. The second are the methods developed by our team during those trial efforts. Workshop materials and manuals describing rigorous, STCW-compliant approaches to developing and conducting assessments and to evaluating simulators to support assessment are included in our reports. Because these materials have been reviewed and tried by representatives of the industry and then

revised as a result of these trials, we believe that they provide a valuable resource to the industry and to the USCG in implementing STCW's mandates. The last are the sample assessments that were developed by our team members and that we offer for review and adoption by those responsible for assessment in their organizations. Our reports and materials will be available on the R&DC website in the coming months: http://www.rdc.uscg.mil.

Dr. Myriam Witkin Smith is a human factors project manager at the Coast Guard Research and Development Center, 1082 Shennecossett Road, Groton, CT 06340-6096. Telephone: (860) 441-2844 or (860) 441-2600

Dr. Marvin C. McCallum is a senior researcher at Battelle Seattle Research Center, 4500 Sand Point Way, N.E., Seattle, WA 98105-0395.

Figure 1. Trial *Lookout* Assessment on board a SeaRiver Maritime ship in San Francisco Bay.

Figure 2. Trial Assessment of *Prepare Main Engine for Operation* on board a SeaRiver Maritime ship in San Francisco Bay