Report No: CG-D-04-97

An Interactive Test of Mariner Competence

U. S. Coast Guard
Research & Development Center
1082 Shennecossett Road
Groton, Connecticut 06340-6096

U.S. Merchant Marine Academy
Department of Marine Transportation
Kings Point, New York 11024-1699

Battelle Seattle Research Center
Seattle, Washington 98105-5428

FINAL REPORT
SEPTEMBER 1996

This document is available to the U.S. public through the
National Technical Information Service, Springfield, Virginia 22161

Prepared for:

U.S. Department of Transportation
United States Coast Guard
National Maritime Center
Arlington, VA 22203-1804

N O T I C E

G.T. Gunther
Commanding Officer
United States Coast Guard
Research & Development Center
1082 Shennecossett Road
Groton, CT 06340-6096

Technical Report Documentation Page

1. Report No. CG-D-04-97	2. Government Accession No. AD-A322240	3. Recipient's Catalog No.
4. Title and Subtitle An Interactive Test of Mariner Competence		5. Report Date September 1996
7. Author(s) M.W. Smith, G.R. Sandberg, M.C. McCallum, R.D. Stewart, D. A. Hard, R.J. Meurn, and D.R. Baker		6. Performing Organization Code
		8. Performing Organization Report No. R&DC 22/96
9. Performing Organization Name and Address U.S. Coast Guard Research and Development Center 1082 Shennecossett Road Groton, Connecticut 06340-6096	U.S. Merchant Marine Academy Department of Marine Transportation Kings Point, N.Y. 11024 Battelle Seattle Research Center 4000 NE 41st Street Seattle, Washington 98105-5428	10. Work Unit No. (TRAIS)
		11. Contract or Grant No.
		13. Type of Report and Period Covered Final Report
12. Sponsoring Agency Name and Address Department of Transportation U.S. Coast Guard National Maritime Center Arlington VA 22203-1804		14. Sponsoring Agency Code National Maritime Center

15. Supplementary Notes

The U. S. Coast Guard technical contact and COTR is Dr. Myriam W. Smith, (860) 441-2844.
The U.S. Merchant Marine Academy contact is Capt. George R. Sandberg, (516) 773-5447.
Captains R.D. Stewart, D.A. Hard, and R.J. Meurn are at the U.S. Merchant Marine Academy, also.
Dr. M.C. McCallum is at Battelle Seattle Research Center; Dr. D.R. Baker is at Univ. of Washington.

16. Abstract

This exploratory study demonstrated the feasibility of interactive testing of mariner competence in the knowledge and application of the Rules of the Road. A desk-top simulator system provides an interactive test that is potentially more accessible and affordable than use of a full-mission simulator. The study also demonstrated the feasibility of automatic scoring of an interactive test. Computerized automatic scoring eliminates the need for an expert examiner and provides objective, repeatable scoring. Administration of the test to sample candidates and analysis of the results demonstrated that multiple-choice items assess classroom knowledge while interactive responses assess the mariner's competence for potential application to bridge operations. Further development of both the test and scoring approaches is required. Recommendations for further development based on experience during the study are provided.

17. Key Words interactive, mariner competence, desk-top simulator, performance measurement, assessment of competence, expert knowledge capture, automated scoring	18. Distribution Statement Document is available to the U.S. public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classif. (of this report)	20. SECURITY CLASSIF. (of this page)	21. No. of Pages	22. Price
UNCLASSIFIED	UNCLASSIFIED		

Form DOT F 1700.7 (8/72) Reproduction of form and completed page is authorized

[BLANK]

PREFACE

This report describes a United States Coast Guard sponsored study exploring the feasibility of using a desk-top simulator for an interactive test of mariner competence. The sample test, designed for a commercially-available desk-top simulator during the study, had as its objectives the assessment of a mariner's understanding of the operational responsibilities imposed on the bridge watchstander by the Rules of the Road. Marine cadets' scores on this test reflected both their operational performance on a full-mission simulator and expert ratings of their demonstrated competence during a broader training program. These findings support the conclusion that the use of the desk-top system was an appropriate approach to meeting the stated objectives. We believe that the findings have generality to the assessment of a variety of mariner competencies.

The study demonstrated that desk-top simulators can make an important contribution to the assessment of mariner competence. Compared to the present paper-and-pencil test, both the expert mariners and the cadets who participated in the study felt that the desk-top simulator-based test would do more to ensure competent mariners. Compared to a full-mission simulator, the desk-top simulator is potentially more accessible and of lower cost. Compared to "practical demonstrations" aboard ship, the desk-top simulator-based test put the study "candidates," cadets soon to be tested for the third mate's license, in a real-time decision-making role that would not otherwise be possible for them.

ACKNOWLEDGMENTS

The authors want to acknowledge and express gratitude for the many people who contributed to the study:

From United States Coast Guard Headquarters, Mr. Stewart Walker, and Mr. Perry Stutman provided guidance and encouragement.

From the United States Merchant Marine Academy, Dr. Warren Mazek, made certain that the full support of the Academy administration was available, and Commanders Ann Sandborn, Paul Zerafa, and Daniel Hunt provided peer review for the test development.

From PC Maritime LTD of the United Kingdom, manufacturers of "Officer of the Watch," the desk-top simulator that we used, Mr. David Edmonds and his staff made timely modifications to the software interface for our test and assisted with the test analysis.

From Houston Marine Training Services, Mr. Brant Houston and Mr. Greg Szczurek gave valuable assistance with the "Officer of the Watch" program. That company is the U.S. representative for PC Maritime Ltd.

[BLANK]

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (CONTINUED)

Page

APPENDICES

LIST OF FIGURES

LIST OF TABLES

AN INTERACTIVE TEST OF MARINER COMPETENCE

EXECUTIVE SUMMARY

INTRODUCTION

The United States Coast Guard (USCG), in fulfillment of its responsibility for marine safety, provides for the examination and licensing of merchant mariners. At the present time, examinations are primarily pencil and paper tests of the candidate's knowledge of a topic. In a time of increased demand for testing of mariner competence, of decreased opportunity for formal on-board training, and of increased acceptance of simulators for training, the USCG sponsored an exploratory study of computer-based interactive testing of mariner competence. The expectation was that an interactive test would allow the examination of a candidate's ability to apply knowledge in a real-time decision-making context, an examination that is potentially better able to predict ability to perform successfully in the operational setting than is the current pencil and paper test. The purpose of our study was to explore the feasibility of interactive testing and its potential benefits, in order to determine whether the approach warranted further, more complete consideration.

OBJECTIVES AND SCOPE OF THE STUDY

The objectives of this study were, first, to explore the feasibility of developing an interactive test, using a desk-top simulator. The use of this type of technology could provide an interactive test that is potentially more accessible and affordable than is a full-mission simulator. Second, we explored the feasibility of automatically scoring the interactive test. Computer-based automatic scoring could free such testing from the requirement for expert examiners and provide objective, repeatable results. Finally, we explored the potential results and benefits of an interactive test. Test performance from a sample of future mariners provided a preview of the benefits to be expected from such assessment.

To limit the study to a workable scope, we explored only the assessment of a candidate's knowledge of, and ability to apply, the Rules of the Road (ROR). To do this, we used an existing desk-top simulator to provide a platform for our tester, an existing USCG Examination Module to define the initial test content, and a sample of United States Merchant Marine Academy (USMMA) cadets as a test population of future mariners. While the scope of the Interactive Rules of the Road Tester (IRORT) Project was limited, the issues considered are far broader and our exploration contributes to a general understanding of assessment of mariner competence by demonstration.

FEASIBILITY OF AN INTERACTIVE TEST

We selected the commercially-available desk-top simulator that best met our requirements for testing knowledge and application of ROR. This simulator presents the user with an "out the window" view which can be rotated 360 degrees and inspected with binoculars, bearing compass, and radar. All traffic ships have appropriate day shapes, lights, and whistle signals for their types. The mariner can control the course and speed of own ship and can sound whistle signals. For our purposes in designing the test, it was important that instructions and multiple-choice questions could be inserted into the scenarios and that a computer record was kept of all mariner and ship actions.

The test content was adapted from a Third Mate's Rules of the Road Examination Module. Our Subject Matter Experts (SME) examined the module and classified items as factual/objective, recognition, or operational. Each type of item was treated differently in the interactive test. Items deemed "factual/objective" by SMEs were presented on the computer in the same multiple-choice format as in the paper and pencil Module. "Recognition" items were treated by allowing the mariner to make observations of traffic ships in a dynamic context before answering embedded multiple-choice questions about them and the possible threat that they presented. "Operational" items, that required an understanding of the responsibilities to act imposed by ROR, were a tiny minority of the items in the paper and pencil Module, but were the focus of the interactive test. Three interactive scenarios were developed, to examine the mariners' ability to apply ROR under daytime, nighttime, and fog conditions.

FEASIBILITY OF AN AUTOMATIC SCORING SYSTEM

We designed an automatic scoring approach that replaced our SMEs' evaluations of mariners' performance with a procedure that could be applied by the computer. To capture the experts' judgments, we used an iterative process of review of each step of the scoring problem, independent input from each one on what was required, and group discussion until a consensus was reached. The agreed-upon basic testing objective was that the mariner be required to demonstrate an understanding of the requirements imposed by navigational law. These were to:

- maintain a good lookout and determine if risk of collision exists
- take appropriate action or maneuver to avoid collision
- determine if own ship's action or maneuver was adequate to avoid collision, and ensure that the action of maneuver does not put own ship in a close quarters situation with other vessels.

After agreeing on these objectives, the SMEs reviewed the three scenarios and selected specific operations that demonstrated understanding of each of these requirements. As an example, maintaining a lookout was demonstrated by appropriate visual search, inspection using binoculars, taking of bearings, and use of radar.

To set performance standards for each of these specific operations, our SMEs proposed the application of two scales: level of "proficiency" or level of "competency." Most of the operations were to be rated as to level of proficiency, which was defined with reference to both navigational law and professional standards as "expert," "qualified," or "unqualified." As an example, a mariner's visual search was expert, qualified, or unqualified as it compared to distributions of percent of time spent looking in each direction that were specified by the SMEs as expected for each scenario. Only a few of the operations were rated for level of competence, which was defined only with reference to navigational law. As an example, the mariner either met the legal requirement to sound a signal at maneuver, or he/she did not.

POTENTIAL RESULTS AND BENEFITS OF INTERACTIVE TESTS

Our final objective was the exploration of the potential results and benefits of an interactive test, compared to a multiple-choice paper and pencil test. Our technical approach to this exploration was to administer the test, designed by our SMEs, to a sample of 100 cadets at the U.S. Merchant Marine Academy. The administration both increased our understanding of the requirements for such a test and gave us a sample of performance data for analysis. For our analysis, we had three types of data for each cadet: performance on our interactive test, performance on a comparison multiple-choice paper and pencil test, and biographical data including grades in relevant courses. In addition, for the 50 First Classmen (seniors) in our sample, we had the grade for a course of full-mission exercises performed on a shiphandling simulator.

Our analysis of the data showed the following:

- The multiple-choice format (paper and pencil or computer-based) provided the highest correlation with our best measure of cadet <u>knowledge</u> -- scores in relevant classroom courses.

- The interactive operational scores provided the highest correlation with our best measure of cadet <u>application-based performance</u> -- scores in the full-mission exercises.

- The <u>combination</u> of multiple-choice and operational scores was needed to provide the highest correlation to the broad-based assessment of cadet capability provided by the full set of cadet biographical measures.

SUMMARY OF TECHNICAL CONCLUSIONS

An interactive Rules of the Road test is feasible, using a low fidelity desk-top simulator. Despite the obvious limitations to the realism of a desk-top system, our test required the cadets to demonstrate the ability to apply knowledge of ROR in a challenging, real-time situation.

Automatic scoring of an interactive test is feasible. Automatic scoring means that administration of the test requires a desk-top simulator and minimum attention from a proctor, but does not require an expert mariner to score each exercise history.

A meaningful assessment of the range of knowledge, skills, and abilities required to successfully fulfill the performance requirements imposed by ROR requires a combination of knowledge-based and application-based components.

SUMMARY OF RECOMMENDATIONS

We recommend further development of an interactive tester using a desk-top simulator for the assessment of ROR competence. Our recommended approach to this development is summarized in the report in Section 5. The approach is sufficiently general to apply to the assessment of other mariner competencies, whether using desk-top simulators, part-task trainers, real equipment, or other settings.

[BLANK]

1.0 INTRODUCTION

1.1 PURPOSE

The United States Coast Guard (USCG), in fulfillment of its responsibility for marine safety, provides for the examination and licensing of merchant mariners. At the present time, examinations are primarily pencil and paper tests of the candidate's knowledge of a topic. In a time of increased demand for testing of mariner competence, of decreased opportunity for formal on-board training, and of increased acceptance of simulators for training, the USCG sponsored an exploratory study of computer-based interactive testing of mariner competence. The expectation was that an interactive test would allow the examination of a candidate's ability to apply knowledge in a real-time decision-making context, an examination that is potentially better able to predict ability to perform successfully in the operational setting than the current pencil and paper test. The purpose of our study was to explore the feasibility of interactive testing and its potential benefits, in order to determine whether the approach warranted further, more complete consideration.

1.2 BACKGROUND

In recent years there has been an increased demand for the assessment of mariner competence by practical demonstration. This demand has been expressed: by the USCG's examination of its Mariner Licensing Program in "Licensing 2000 and Beyond;" by a recent revision of mandates in the International Maritime Organization's Standards for Training, Certification, and Watchkeeping (IMO STCW); by legislative requirements in the Oil Pollution Act of 1990; and by a recent USCG-sponsored study by the National Research Council entitled, "Simulated Voyages." (See United States Coast Guard, 1993a; International Maritime Organization, 1995; Oil Pollution Act of 1990; and National Research Council, 1996, respectively.) Implementation of the mandates and recommendations of these efforts requires considerable planning and investigation. Our study is intended as a contribution to that process.

The USCG sponsored a previous study of the feasibility of interactive assessment (Mariner Licensing Device, 1987; Gardenier, Flyntz, Spears, Willis, and North, 1987). The project developed a prototype "Mariner Licensing Device" that could be used to test for a variety of watchstanding, shiphandling, and Rules of the Road competencies. The device was a low-fidelity microprocessor-based simulator fitted with a wheel and a minimum of controls and indicators. Study candidates drove the "ship" through planned scenarios, while the processor maintained a record of transit performance that was to be scored later by an expert mariner. This approach was not developed further because of the high cost of the customized device and because of the cost and difficulties of scoring by an expert. Our approach was intended to overcome these disadvantages. Rather than a customized device, our study used commercially-available simulator software, intended to run on an off-the-shelf personal computer (PC). Rather than relying on an expert scorer, a major component of our study was the exploration of the complex, technical problem of automatically scoring interactive performance.

1.3 OBJECTIVES AND SCOPE OF THE STUDY

The intent of this study was to explore the feasibility and potential benefits of using interactive simulation for testing mariner competence. Our study explored:

1. the feasibility of developing an interactive test, using a desk-top simulator. The use of this type of technology could provide an interactive test that is more accessible and affordable than is a full-mission simulator.

2. the feasibility of automatically scoring an interactive test. Computer-based automatic scoring could free such testing from the requirement for expert examiners and provide objective, repeatable results.

3. the potential results and benefits of an interactive test. Test performance from a sample of future mariners provided a preview of the benefits to be expected from such assessment.

To limit the study to a workable scope, we:

- explored only the assessment of a candidate's knowledge of, and ability to apply, the Rules of the Road (ROR) (United States Coast Guard, 1996)

- used an existing commercially-available desk-top simulator to provide a platform for our tester

- used an existing USCG Examination Module to define the test content

- used United States Merchant Marine Academy (USMMA) cadets as an accessible population of future mariners

While the scope of the Interactive Rules of the Road Tester (IRORT) Project was limited, the issues considered are far broader and our exploration contributes to a general understanding of assessment of mariner competence by demonstration.

1.4 THIS REPORT AND ASSOCIATED TECHNICAL DOCUMENTS

This report is the final report on the study. As an overview, the report is organized as follows:

SECTION 2.0 FEASIBILITY OF AN INTERACTIVE TEST corresponds to Objective 1 above and describes the initial development of the test. This section is supported by Appendix A which summarizes software capabilities recommended for an interactive tester, by Appendix B which summarizes the specific testing objectives of our interactive test, and by Appendix C which summarizes the initial conditions for our test scenarios.

SECTION 3.0 FEASIBILITY OF AN AUTOMATIC SCORING SYSTEM corresponds to Objective 2 above and describes the development of the automatic scoring approach. A summary of our interactive performance measures is included in Appendix C.

SECTION 4.0 POTENTIAL RESULTS AND BENEFITS OF AN INTERACTIVE TEST corresponds to Objective 3 and describes the administration of our test to a sample of future mariners and an analysis of the results.

SECTION 5.0 TECHNICAL CONCLUSIONS AND RECOMMENDATIONS provides a summary of lessons learned for future developers or evaluators of interactive tests.

SECTION 6.0 IMPLEMENTATION ISSUES reports implementation issues that we identified during the study for further consideration.

SECTION 7.0 SUMMARY OF RECOMMENDATIONS provides a brief overview of our recommendations.

SECTION 8.0 A RELATED STUDY OF ASSESSMENT OF COMPETENCE provides a brief preview of another project which will provide a broader study of some of the issues treated here.

A number of other reports on the study were prepared. These included a preliminary report on the issues associated with the initial development of the IRORT (Stewart, Sandberg, Meurn, and Hard, 1994); a shorter, less detailed version of the present report (United States Coast Guard, 1996); and a number of short papers available in conference proceedings and periodicals (Sandberg, Stewart, Smith, McCallum, 1996; McCallum, Smith, Sandberg, Hard, Meurn, and Stewart, 1996; Sandberg and Stewart, 1995/6; and McCallum, Smith, Sandberg, Hard, and Stewart, 1995).

2.0 FEASIBILITY OF AN INTERACTIVE TEST

2.1 EXAMINATION IN RULES OF THE ROAD

We selected ROR competence as the subject matter both because of the regulatory requirements for examination in this subject and because of the importance of its application to watchstanding and to marine safety. Regulations require examination in this subject for all the major deck licenses (Code of Federal Regulations, 1994). The Oil Pollution Research and Technology Plan prepared under the authority of Title VII, Oil Pollution Act of 1990 calls for the development of interactive testing for ROR. The Amendments to IMO's STCW called for demonstration of competence in watchkeeping, including a "Thorough knowledge of the content, application and intent of the International Regulations for Preventing Collisions at Sea" (International Maritime Organization, 1995). As early as 1970, when simulation was in its infancy, a study of the examination process recommended that the USCG investigate the possibility of testing in ROR by "simulated situations," rather than by written test (Jensen and Shimberg, 1970).

2.2 SELECTION OF APPROPRIATE TECHNOLOGY

The first objective of the study was to explore the capability of desk-top simulation to provide an interactive environment for the examination of some aspects of mariner competence, an environment with more realism than a paper and pencil test but with less potential expense than a full-mission simulator. From among the commercially-available desk-top systems, we selected PC Maritime Limited's "Officer of the Watch" (OOW), as most representative of the capability required (PC Maritime Limited, 1993a and 1993b; Hughes, 1993). The selection of OOW is not intended as an endorsement of this product but, because it is intended by its manufacturer to train ROR, it does have many of the features needed for testing ROR. During the development and evaluation of the interactive test, we made new demands on the software that it was not designed to meet. The resulting identification and recommendation of additional features required for an interactive test, listed in Appendix A, is not meant as criticism of this training software.

OOW consists of a ship "simulator" which runs as software on a PC, with the keyboard and mouse as controls. The simulator presents the user an "out-the-bridge-window" view, which can be rotated 360 degrees and inspected with binoculars, bearing compass, and radar. Day, night, and restricted visibility are possible. All traffic ships have appropriate day shapes, lights, and whistle signals for their types and circumstances. The "mariner" can control the course and speed of own ship and can sound appropriate whistle signals. For our purposes, it was important that instructions and multiple-choice questions could be inserted into the scenario and that the software kept a log of all user actions and responses and a "playback reel" of all ship actions. A capability critical to our study was the separate "Course Designer" software, which was intended to allow an instructor to design custom scenarios. We developed the IRORT scenarios using the Course Designer.

2.3 DESIGN OF THE INTERACTIVE SCENARIOS

A single Third Mate's Rules of the Road Examination Module in a paper multiple-choice format was provided by the USCG (Stewart, Sandberg, Meurn, and Hard, 1994; United States Coast Guard, 1993b). The Module was used as the source of testing objectives for the IRORT, both to restrict the study to a manageable scope and to ensure that the interactive test was comparable to the current examination process. Our subject matter experts (SMEs) analyzed the test items and identified three types: 1.) factual/objective items on definitions, technical information, etc.; 2.) recognition items on vessel types from day shapes, lights, or sounds; and 3.) operational/action questions that require a comprehension of responsibilities imposed and actions required by the traffic situation. The recognition and operational/action questions were analyzed further to develop the specific testing objectives that would guide the scenario design. These objectives are described more specifically in Appendix B.

Each type of item was treated differently in the design of the scenarios. The simplest requirements, to know definitions, appeared both as multiple-choice questions and as prerequisites for appropriate action in the scenarios. Recognition of traffic ships, and of any threat they might represent, was based on observing them over several minutes, using all available means (visual search, binoculars, bearing compass, and radar) before answering embedded multiple-choice questions about them. (As an example, "The vessel on your port bow is a _______," with four alternatives following.) This recognition is presumably closer to real world conditions than the recognition from the paper illustrations presently used in testing. The operational/action questions, those requiring the most complex type of understanding, were a minority of the questions in the USCG Module but were the focus of the IRORT scenarios. The mariner was required to demonstrate the actions required of him/her by the ROR. (For example, keep a good lookout.) This demonstration is presumably closer to real world operations than is the selection of the appropriate action in a multiple-choice item. For further descriptions of the scenario development, see Sandberg and Stewart (1995/6) and Stewart, Sandberg, Meurn, and Hard (1994).

The IRORT was designed in five parts, each approximately 20 minutes in length. The first part was a customized familiarization exercise that allowed the test takers to examine all the system features needed for the test and practice the control actions that would be available to him/her. The next part was not interactive but consisted of 14 factual/objective items from the USCG Module presented on the computer screen in the same multiple-choice format as in the original paper version. Three truly interactive scenarios followed. The first was a daytime scenario that began by placing the mariner on the bridge, able to assess a dynamic traffic situation by visual search, binoculars, bearing compass, and radar but without control of his/her vessel. After allowing the mariner a set period of time to assess the situation, IRORT presented a series of embedded multiple-choice questions, asking for the recognition of the several traffic ships and for the identification of a possible threat of collision from one of them. After the recognition questions, the mariner was given control of own ship's course and speed and was required to demonstrate the appropriate actions to avoid collision in the scenario. A nighttime scenario followed, requiring the recognition of traffic vessels from their lights and a demonstration of appropriate actions in a second situation. The last was a "fog" scenario which required the recognition of whistle signals and the demonstration of the special requirements imposed by the ROR in restricted visibility. A number of expert mariners who had not been involved in the development took the "test" and provided initial peer review. A more specific description of the content of the each of the interactive scenarios is provided in Appendix C.

3.0 FEASIBILITY OF AN AUTOMATIC SCORING SYSTEM

3.1 THE SCORING PROBLEM

The second objective of the study was to demonstrate the feasibility of scoring the interactive test results automatically by the computer. Scoring of the multiple choice items, both the factual items and the embedded recognition items, was relatively straight-forward. The scoring of operational actions performed by the test takers during the interactive scenarios was a far more complex matter. The desk-top simulator software recorded a variety of measures, but the problem remained to specify what constituted acceptable performance, whether a test taker had "passed." The usual approach to evaluating performance during simulator training or testing has been judgment by expert observer (National Research Council, 1996). What we wanted was to reduce the judgments of our SMEs to a reliable procedure that could be applied to the recorded performance by computer. To capture the expert judgments, we used a modification of the "Delphi method" (Meister, 1985). Basically, this is an iterative process of review, independent input, and group discussion until a consensus is reached. Our SMEs agreed on each step described below before we went on to the next. For another discussion of our development of the scoring approach, see McCallum, Smith, Sandberg, Hard, Meurn, and Stewart (1996).

3.2 DEVELOPMENT OF THE SCORING APPROACH

3.2.1 Basic Testing Objectives and Operational Measures

For our SMEs, the basic testing objectives were the demonstration of an understanding of the requirements imposed on the mariner by navigational law. First, they identified three basic requirements of navigational law that are universally applied during a watch. Then, they reviewed the interactive scenarios and identified the specific actions required from the mariner that could represent understanding of each of the three requirements. The legal requirements and examples of representative actions follow:

- Maintain a good lookout and determine if risk of collision exists was represented by:
  - visual search in all directions
  - radar viewing
  - binocular viewing of traffic ships
  - visual bearings taken on traffic ships

- Take appropriate action or maneuver to avoid collision was represented by:
  - timely sounding of the correct signal at the time of maneuver
  - sounding of fog signal in the fog scenario
  - reduction of ship speed in the fog scenario following radar loss

  (Our SMEs intended to include the action taken to change speed or direction and the adequacy of the projected closest point of approach (CPA) with the threat vessel at the time of the action. Unfortunately, we were unable to extract this important measure from the exercise record.)

- Determine if own ship's action or maneuver was adequate to avoid collision, and ensure that the action or maneuver does not put own ship in a close quarters situation with other vessels was represented by:
  - minimum actual CPA with any vessel
  - taking of new visual bearings on traffic ships
  - new radar viewing

3.2.2 Performance Standards for the Operational Measures

The next step was the establishment of performance standards for the operational measures. The SMEs determined that the standard of performance they would expect for each representative measure could be specified in terms of "proficiency" levels or "competency" levels. These concepts and their levels were defined as follows:

- Proficiency Level is defined with reference to both navigation law and professional standards. Considering both, it is the consistency of performance with legally mandated actions, as defined by navigation law; and with the indicated level of prudent seamanship (expert, qualified, unqualified) in the operational application of navigation law. These three proficiency levels are:

  - Expert: Performance is fully consistent with all legal mandates and meets the highest professional standards of prudent seamanship in the operational application of navigation law.

  - Qualified: Performance is fully consistent with all legal mandates and meets acceptable professional requirements of prudent seamanship in the operational application of navigation law.

  - Unqualified: Does not meet one or both of the legally mandated actions and/or acceptable professional requirements of prudent seamanship in the operational application of navigation law.

- Competency Level is defined with reference to navigation law only. It is the consistency of performance with legally mandated actions, as defined by navigation law. The two levels of competency are:

  - Competent: Performance is fully consistent with legally mandated actions.

  - Incompetent: Performance is inconsistent with legally mandated actions.

The SMEs determined that most of the measures were appropriately scored for a level of "proficiency," as "expert," "qualified," or "unqualified." The next step, and the most difficult, was the determination of standards of performance for each of the representative measures: that is, what constitutes "expert," etc. Performance standards were recommended and reviewed until a consensus was obtained for each measure in each scenario. The SMEs specified separate standards for each scenario, considering the conditions -- weather, behavior of traffic ships, own ship size and speed, etc. -- as they would have if they had actually been evaluating the mariner's performance. The separate standards for scenario conditions are an important factor in the validity of the automatic scoring procedures.

As an example of a performance standard, Table 1 illustrates the results of this process for visual search during the first six minutes of the daytime scenario. The table provides the ranges of percentages of the total visual search time in each direction that the team of SMEs judged to correspond to each proficiency level for the conditions in that scenario. Only a very few measures were selected by the SMEs to be scored for "competency." An example of these is the sounding of the correct signal at the time of maneuver: either it is sounded or it is not. The complete set of operational action measures in each scenario and the standards established for each are provided in Appendix C.

Table 1. An Example of Performance Standard for One Operational Action Measure

Percentage of Total Visual Search Time in Each Direction During First Six Minutes				**Performance Level**
Forward	**Starboard**	**Aft**	**Port**	
35-50	*30-45*	*5-15*	*5-15*	Expert
25-80	*15-80*	*5-35*	*5-35*	Qualified
Neither Expert nor Qualified				Unqualified

3.3 IMPLEMENTATION OF THE SCORING SYSTEM

To meet the project objective of demonstrating the feasibility of automatic scoring, it still remained to translate our performance standards into specific scoring procedures that could be applied to the exercise records produced by the desk-top simulator software. For some desired measures, the recorded actions could be extracted and grouped by computer. In other cases, clerical involvement was needed. For some measures (for example, minimal CPA in a scenario) more knowledgeable involvement was needed to extract the results. While we did not have the necessary resources to develop automatic means of scoring all the desired measures, we believe that it would have been technically possible to develop an entirely automatic process. Our experience with the scoring contributed to our list of recommended software features listed in Appendix A.

Numerical scores had to be assigned for each measure, based on comparison to the SMEs' performance standards. For this study, a relatively arbitrary convention was adopted of awarding two points for each instance of proficient action, one point for each instance of qualified or competent action, and zero points for each instance of unqualified or incompetent action for all the measures. As an example, the standards for the visual search, illustrated in Table 1, above would have been applied by awarding two points for an exercise record that showed the distribution for "expert" performance, one point for one that showed "qualified" performance, and zero points for performance which did not meet those standards. For most measures, there were multiple opportunities to gain points over the three interactive scenarios. The final score for a measure was the total points awarded over multiple opportunities. (Other conventions would have been possible, including weighting schemes with more points awarded for actions deemed more important by the SMEs or points subtracted for especially egregious omissions.) The operation measures for which final scores were achieved are summarized in Table 2.

Table 2. Summary of Operation Measures for Which Scores Were Available

Performance Measure	**Description**
Visual Search	Score based on distribution of time spent visually searching each of four quadrants of view during the scenario observation period. Points totaled over scenarios.
Binocular Viewing	Score based on frequency of viewing each vessel through the binoculars during the scenario observation period. Points totaled over scenarios.
Visual Bearing	Score based on frequency of visual bearings on each vessel taken during the scenario observation period. Points totaled over scenarios.
Radar Viewing	Score based on frequency and total duration of radar viewings during the scenario observation period. Points totaled over scenarios.
Maneuvering Signal	Proportion of times a correct signal was sounded within 30 seconds of a maneuver
Minimum CPA	Score based on minimum closest point of approach to other vessels throughout each test scenario. Points totaled over scenarios.
Radar Maneuver	Score based on frequency and total duration of radar viewings during the first six minutes after the first own ship maneuver. Points totaled over scenarios.
Fog Signal	Score based on time of first sounding of the restricted visibility sound signal during the fog scenario
Fog Speed	Score based on time and level of action to reduce speed following radar failure during the fog scenario

4.0 POTENTIAL RESULTS AND BENEFITS OF AN INTERACTIVE TEST

4.1 OBJECTIVE AND APPROACH

Our third and final objective in this feasibility study was to explore the potential results and benefits of an interactive test, compared to a paper and pencil test. Our approach was to administer the IRORT to cadets at the USMMA. This population was accessible to us and provided a relatively-complete set of biographical information. Actually administering the test gave us a greater understanding of the requirements for desk-top system capabilities. This understanding is reflected in Appendix A, which lists the system capabilities recommended. In addition, actually administering the test gave us a set of performance data for an analysis that would contribute to an understanding of the potential results and benefits of an interactive test. See McCallum, Smith, Sandberg, Hard, Meurn, and Stewart (1996) and McCallum, Smith, Sandberg, Hard, and Stewart (1995) for additional descriptions of our evaluation of test results.

4.2 ADMINISTRATION TO U.S. MERCHANT MARINE ACADEMY CADETS

We administered the IRORT to 100 cadets at the USMMA. We included 50 First Classmen (seniors), 25 Second Classmen (juniors), and 25 Third Classmen (sophomores). The cadets were all volunteers from programs of study leading to a Third's Mate's Unlimited License who were paid a small fee for participating. All were computer literate, making routine use of computers in their school work.

The cadets were tested in groups of three to six in a computer laboratory for two evenings each. They were briefed on the general purpose of the study and, then, each signed a consent form permitting the use of his/her performance data and Academy grades for the study. One half took the IRORT on his/her first evening and the other half took a paper and pencil test consisting of selected items from the USCG question bank for comparison. The cadets taking the IRORT on a given night were given five "scenarios" -- familiarization, factual multiple-choice, daytime, nighttime, and fog -- and finished the entire session in under two and a half hours. Those taking the pencil and paper test were allowed the same amount of time but generally finished in less than an hour. Each one returned on a second evening to take the second test and to complete a questionnaire, which asked for a variety of biographical information.

4.3 DATA ANALYSIS AND SELECTED FINDINGS

4.3.1 Data Available for Analysis

We had three types of data available for our analyses:

1. Biographical data on the cadets included existing course grades and questionnaire responses. To represent the cadets' relevant knowledge, existing grades were recorded from a number of courses that the faculty selected as relevant to ROR. The best available approximation to an independent assessment of a cadet's ability to perform in an operational setting was a grade, available only for the First Class cadets, from a course that used the full-mission shiphandling

simulator at the USMMA. This course is designed to teach and practice passage planning, teamwork, and bridge operations during realistic port arrivals and departures. ROR are included but not emphasized. Faculty observers of the exercises in this course use a structured checklist of performance requirements to provide a grade (Meurn, 1995, 1990). Because of the relative difficulty of obtaining appropriate samples of shipboard performance to use as independent assessments, there is a history in the marine industry of the use of performance on a full-mission simulator to fill this role (Moynehan, Hanley, and Pittsley, 1985; Hanley, 1984; Hammell, Gynther, Grasso, and Gaffney, 1981.) Questionnaires administered during the study documented the cadets' year and their self-ratings of knowledge of ROR as obtained from classroom, simulator training, and sea projects. USMMA faculty provided ratings on their perceptions of the relevant knowledge and skills of each cadet.

2. Responses on a set of USCG multiple-choice paper and pencil items. The comparison paper and pencil multiple-choice score that we used in the analysis was a sample of items that were selected by our SMEs as presenting information and situations similar in intent and difficulty to those presented in the interactive scenarios.

3. Responses on the IRORT. These included a small sample of USCG factual/definition multiple-choice items presented on the screen and IRORT recognition multiple-choice items embedded in the interactive scenarios. In addition, IRORT responses included the operational actions summarized in Table 2 in Section 3.4. These last action scores were the focus of the analysis: do such items add significantly to the effectiveness of the testing process?

4.3.2 Test Validity

Ideally, we would like to have a measure of a mariner's operational performance when we decide whether he/she should have a license. Since this is impractical, we substitute a number of different tests (along with other requirements). The first question to ask of any test is the extent of its "validity," that is, the extent to which an individual score on that test can support the intended inferences (Gatewood and Feild, 1994; Anastasi, 1988; and Berk, 1984). We want to be able to infer that IRORT accurately measures an individual's knowledge of ROR, and his/her ability to apply ROR in operational performance. In addition, we have a second question to ask, whether we infer from the IRORT score something different than that which can be inferred from the USCG multiple-choice pencil and paper test.

One of the principal strategies for demonstrating the validity of a test is by examining its "content" validity, the extent to which a test samples the content of the tasks for which the individual is being evaluated. This type of validity is considered by the testing industry to be especially critical, even sufficient, for licensure examination (Gatewood and Feild, 1994; Kane, 1982; Shimberg, 1981). We can ask, first, whether the IRORT samples the test taker's knowledge of ROR. IRORT does sample the same knowledge content as does the USCG multiple-choice pencil and paper test, a similarity that was ensured when the initial IRORT test objectives were developed by an examination of the pencil and paper test. (See Section 2.3 and Appendix B for descriptions of that development.)

Our SMEs developed additional content to be sampled by the IRORT: that is, the understanding of the responsibilities imposed by ROR in a real-time situation. (Development of this additional content is described in Section 3.2.1 and Appendix C.) We can illustrate the substantial difference in what is required of the test taker by the multiple-choice and interactive test formats with an actual example. To test for an understanding of Rule 8, Action to Avoid Collision, the USCG multiple-choice pencil and paper module contained one multiple-choice item asking for a recognition of the requirement to take action. In contrast, this Rule was represented in our IRORT by the requirements in three different scenarios for the test taker to:

- sound the correct signal at time of maneuver
- avoid a close quarters situation with all other vessels in the scenario
- ascertain whether the action taken had the desired effect by taking visual bearings and viewing the radar for an adequate time after the completion of the maneuver.

This example illustrates an important point. The interactive approach samples a type of content that is not sampled by the paper and pencil multiple-choice test. *The interactive test is not merely an alternative to the multiple-choice test for testing mariner knowledge. If the objective of testing is to evaluate the ability to apply knowledge, rather than knowledge alone, an interactive approach is essential.*

Another strategy for evaluating a test is an examination of its “criterion” validity, that is, the extent to which a score on the test corresponds to an independent criterion, or measure, of the knowledge or ability in question. The simplest approach to criterion validation is to examine the ability of the test to discriminate the level of expertise, “expert” or “novice,” of the test takers (Kelly, 1988; Vreuls and Obermayer, 1985; and Berk, 1984). In our context, we did find that the First Classmen performed better on the average than the underclassmen on a variety of IRORT performance measures. However, the sample pencil and paper multiple-choice test also did well in discriminating the First Classmen, who were only a few months away from taking the USCG licensing examinations, from the others. Therefore, this relatively simple analysis of criterion validity did not demonstrate any advantage of the interactive test over the easier-to-administer paper and pencil test.

The most stringent and ambitious approach to evaluating a test is to examine its criterion validity by comparing the scores in question to an independent measure of the “criterion” performance. In other words, how much does the score on the test being evaluated tell us about an individual’s potential performance on the operations of interest? Such an evaluation is not often done in competency testing because of the difficulty in finding an appropriate independent measure of the criterion performance. In our study, we had the variety of biographical data from the participating cadets as potential measures of their abilities. For an independent measure of the participating cadets’ knowledge of ROR, we used selected course scores. To determine whether the IRORT could measure something different from that measured by the pencil and paper test, we required an operational performance sample to provide an appropriate independent criterion measure. The best independent measure we had of the cadets’ ability to apply the knowledge learned at the Academy was the First Classmen's scores in the full-mission simulator course (described in Section 4.3.1). The results of our analysis are described below.

4.3.3 Differences in What the Formats Measured

It is widely believed that a mariner's competence in the marine environment could be better assessed in real-time decision-making situations than with a static, paper and pencil multiple-choice test. Paper and pencil tests measure primarily classroom knowledge, and not necessarily an individual's ability to apply that knowledge in an operational situation. The purpose of competency testing for mariners is to provide a measure of how well a mariner will perform in the marine environment and to ensure that only competent individuals are licensed to operate vessels. Our study was designed to determine whether IRORT might indeed provide a better measure of a mariner's ability to perform in the marine environment than did the existing multiple-choice format. We compared cadet scores in each format to those biographical measures that provided the best independent measures of knowledge and of their ability to apply that knowledge.

Our analysis (SPSS Inc., 1990; Harris, 1975) showed that knowledge was best represented by the standardized grade for a cadet's most recently-taken ROR-relevant course. This grade was compared by correlational techniques to each of three sets of scores that had been generated during the study: scores for the sample of USCG pencil and paper multiple-choice items, for the IRORT's factual/definition and recognition items that had been presented on the computer as multiple-choice items, and for the IRORT's action items summarized in Table 2. The results of this analysis are illustrated in Figure 1. Both the pencil and paper and IRORT multiple-choice scores had moderate, significant correlations with this grade, as shown by the first two bars in the figure. The IRORT action score did not correlate with this grade, as shown by the last bar. This pattern shows that the multiple-choice formats, whether by paper and pencil or by computer, were testing classroom knowledge, while the IRORT action measures were not testing the same thing.

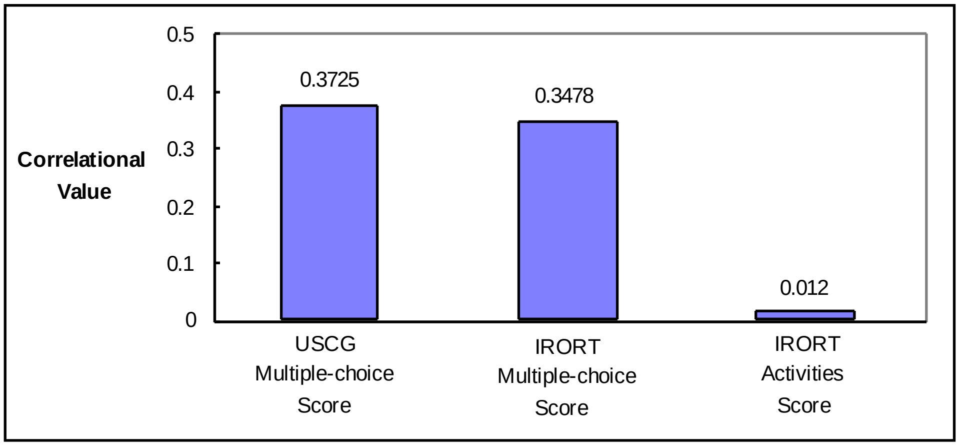

Figure 1. Correlation of the Study's Test Scores with Selected USMMA Course Grade

The grade from the full-mission simulator course taken by the First Classmen gave us our best independent measure of the ability to apply ROR. While that course was not designed specifically to assess ability to apply ROR, the bridge watchstanding task did include many of the same actions as those that were scored for the IRORT. We compared the simulator course grade with each of the three sets of scores as before: for the sample of USCG pencil and paper multiple-choice items, for the IRORT factual/definition and recognition items that had been presented on the computer as multiple-choice items, and for the IRORT's action items summarized in Table 2. Our analysis found that both the paper and pencil and IRORT multiple-choice test formats had low, non-significant, negative correlations with the cadets' simulator grades, indicating that the simulator grade and the two multiple-choice test items measured different aspects of competence. These relationships are illustrated by the first two bars in Figure 2. In contrast, we found a moderate, significant, positive correlation between the simulator grades and IRORT action scores, as illustrated by the last bar in Figure 2. This last correlation demonstrates that the IRORT action score provides a moderate prediction of the simulator score. *The two multiple-choice formats do not capture the important ability to apply ROR. However, the IRORT action score provides a better measure of the ability to apply ROR and, therefore, it is potentially a better predictor of how a cadet might perform on the bridge of a ship.*

Our results show that while classroom knowledge may be adequately measured with paper and pencil multiple-choice tests, measurement of the ability to apply that knowledge requires the use of more interactive approaches. The USCG's current testing approach of relying solely on paper and pencil tests ensures knowledgeable mariners, but not necessarily mariners with the appropriate operational skills.

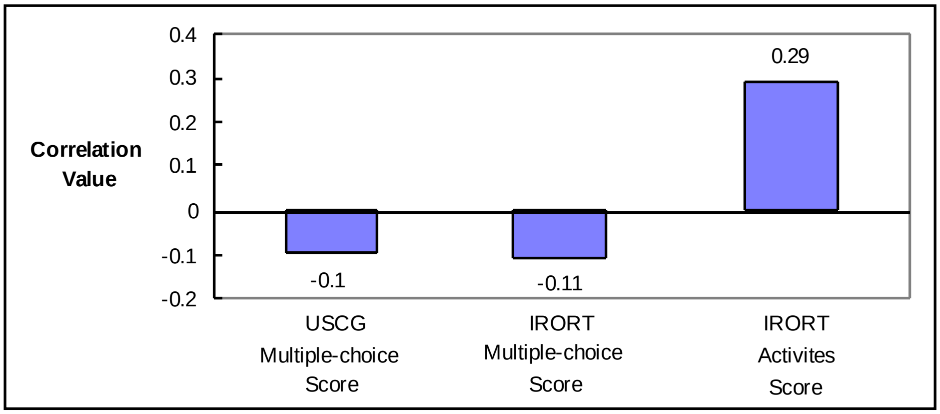

Figure 2. Correlation of the Study's Test Scores with Full-mission Simulator Grades

4.3.4 A Broad Test of Knowledge and Application

The IMO STCW (International Maritime Organization, 1995) code requires that a license candidate demonstrate a "thorough knowledge of the content, application and intent of the International Regulations for Preventing Collisions at Sea." This requirement implies that knowledge and the ability to apply it must be broadly assessed. We used the full set of available biographical measures (described in Section 4.3.1) as the criterion for what a broad assessment should measure. This analysis is summarized in Figure 3. The IRORT multiple-choice component showed a moderate, significant correlation with the biographical measures, as shown by the first bar in this figure. The combination of the IRORT multiple-choice and action measures together correlated more highly with the biographical measures, as shown by the second bar. Further refinement of a complete test of mariner competence, beyond what we attempted in this feasibility study, would weight individual test items/components on the basis of their predictive value. We further explored our data to identify potential weightings that would further increase these correlations. These analyses resulted in an even higher correlation, as shown by the last bar in Figure 3. This last result shows the need for the testing of both knowledge and ability to apply knowledge for a complete test of mariner competence. In addition, the analyses summarized in Figure 3 demonstrate the potential, future value of item weightings.

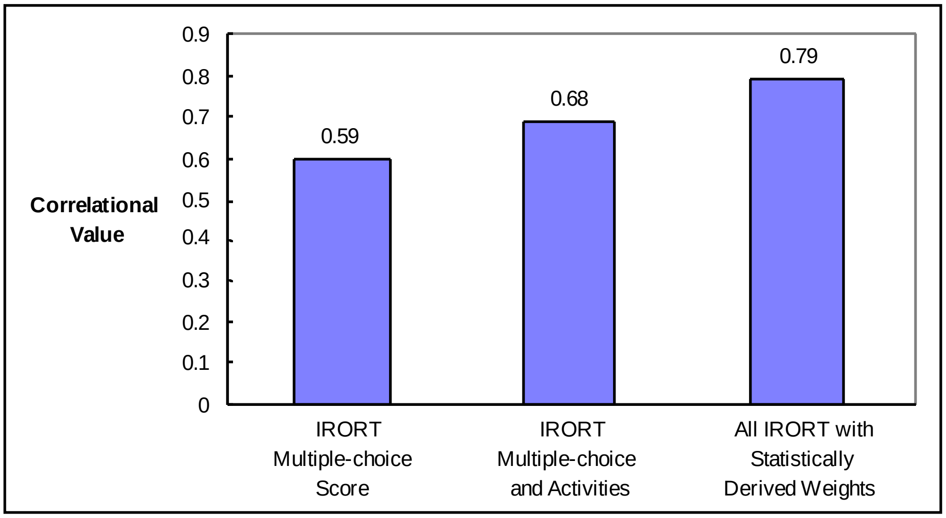

Figure 3. Correlation of IRORT Scores with Cadets' Biographical Data

5.0 TECHNICAL CONCLUSIONS AND RECOMMENDATIONS

5.1 FEASIBILITY OF AN INTERACTIVE TEST

An interactive Rules of the Road test is feasible, using a low fidelity desk-top simulator. Despite the obvious limitations to the realism of a desk-top system, our test required the test taker to demonstrate the ability to apply knowledge of ROR in a challenging, real-time situation. We strongly recommend further development of the concept. We recommend that development include the following components.

1. Specification of the testing requirements for ROR testing. A comprehensive list of the knowledge, skills, and abilities required of a competent bridge watch officer that are appropriately considered "Rules of the Road" is needed. The list should include both knowledge-based and application-based items. This specification would guide the design of the actual test and the selection of appropriate performance measures, and would serve as a basis for the evaluation of testing effectiveness. The testing requirements on which we based our ROR test are documented in Appendix B here to provide an example to future developers.

2. Design of test scenarios to meet the specifications. Implementation would require a pool of multiple, equivalent scenarios to ensure that candidates prepare broadly for the test rather than "memorize" a very few scenarios. The scenarios we developed for our test are documented in Appendix C.

3. Development of interactive performance measures. Additional performance measures are needed to maximize the sensitivity of the test in assessing mariner competence. In addition, scoring criteria need to be established for each measure. Our development of performance measures is discussed in Section 3 and Appendix C.

4. Development of new desk-top simulator features. During our study we identified features that are essential or desirable for a desk-top simulator-based test. We have listed these in Appendix A.

5.2 FEASIBILITY OF AN AUTOMATIC SCORING SYSTEM

Automatic scoring of an interactive test is feasible. Automatic scoring means that administration of the test requires a desk-top simulator and minimum attention from a proctor, but does not require an expert mariner to score each exercise history. Automatic scoring would be more reliable than that done by an expert: that is, a given response would always be scored the same way, without the variability possible among human judges. In addition, computer-based scoring offers the potential for calculating a candidate's grade as the test is being taken so that a record could be provided at the end of the test. Validity of the approach is assured by separate standards for each scenario, allowing the consideration of the factors that an expert would consider in evaluating performance.

We recommend the general approach that we took to capture the judgments of expert mariners as to how they would have scored individual candidate performance on the test for the given scenario conditions. Our SMEs specified:

- the responsibilities imposed by ROR on a navigational watch:
  - to maintain a good lookout and determine if risk of collision exists
  - to take appropriate action or maneuver to avoid collision
  - to determine if own ship's action or maneuver was adequate to avoid collision, and ensure that the action of maneuver does not put own ship in a close quarters situation with other vessels

- the observable actions that would demonstrate the understanding and fulfillment of these responsibilities. The specific actions depend on the content of a specific scenario (see Section 3.1). The actions scored in our test are documented in Appendix C.

- the timing, frequencies, or levels of these operations that would constitute adequate performance. These also depend on the content of a specific scenario. Examples appear in Section 3.2 and in Appendix C.

We recommend further exploration of two scoring issues which we considered only briefly:

- the design of a procedure for weighting specific test items in the calculation of the total test score. Both the specification of weights by SMEs and statistical techniques should be included.

- the specification of the cutoff score for passing or failing the test

We also recommend:

- the exploration of the possibility that a computerized system might provide a report of a candidate's strengths and weaknesses as a guide to his/her future study

The actual mechanisms of collecting and manipulating performance data will be specific to a particular scenario and a particular desk-top system and are not dealt with in any detail here. While many of the features we required were not available with the commercial software, we believe that they are well within the capabilities of current technology. See Appendix A for our recommendations.

5.3 POTENTIAL RESULTS AND BENEFITS OF AN INTERACTIVE TEST

A meaningful assessment of the range of knowledge, skills, and abilities required to successfully fulfill the performance requirements imposed by ROR requires a combination of knowledge-based and application-based components. Based on our findings and our experience during the study, we recommend that the USCG encourage further efforts to develop testing by application-based interactive means and, eventually, require such assessment. How knowledge-based and application-based elements can best be combined was not considered in our study. Our recommendation, that the USCG encourage the development of application-based approaches applies not only to assessment of ROR competence, but also to assessment of other mariner competencies now assessed with paper and pencil multiple-choice tests.

6.0 IMPLEMENTATION ISSUES

6.1 RESPONSIBILITY FOR AN INTERACTIVE TEST

An issue of frequent discussion during the study was whether an interactive test would be developed by the USCG or under USCG sponsorship, or whether such a test might be developed by a third party and offered for USCG approval. The technical issues discussed in Section 5.0 are the same, whoever develops the test.

A related issue was the proper balance between the industry's desire to know the contents of a licensing examination and the security required to discourage cheating . For the present paper and pencil test, there are a very large number of questions and these questions, with their answers, are available to the public for review and study. The need for security is met by random selection from the large set for a specific module and the guarding and occasional replacement of the module. This same approach will not be practical for an interactive test, which will never include a comparably-large number of alternative scenarios. If an interactive test is developed by a third party, the proprietary interests of that party will need to be protected, in addition to protecting the security of a licensing examination. New mechanisms for industry review and acceptance and for security must be considered.

6.2 ADDITIONAL IMPLEMENTATION AND ADMINISTRATIVE ISSUES

During the study, a variety of issues were identified that would require resolution before implementation of an IRORT. The most important are included here for further consideration.

1. <u>Cost of development and administration</u>. The cost of the development of an interactive examination would include not only the cost of the software but also the cost of developing the test and the automatic scoring, as described in Section 5. In addition, the cost of its administration would be considerably greater than the individual mariner is accustomed to paying for examinations. We join the National Research Council (1996) and the Licensing 2000 focus group (United States Coast Guard, 1993a) in identifying cost as a major issue in the move to demonstrations of mariner competence.

2. <u>Trade-off between interactive tests and paper and pencil modules</u>. If a candidate does take an interactive test, could he/she be excused from taking a corresponding paper and pencil test module? We did not consider this issue directly. However, our methodology for designing interactive tests suggests an approach for determining appropriate trade-offs. If a list of the knowledge, skills, and abilities required of a competent watch officer that are appropriately considered "Rules of the Road" is available, entries on that list can be assigned to interactive or multiple-choice formats. Multiple-choice items can be presented on the computer or by paper and pencil. The appropriate implementation becomes "how can all the required knowledge, skills, and abilities be sampled validly and practically by a combination of formats?"

3. Effects of an interactive test on training. The type of test that will be faced by the license candidate has major implications for the training process. At the present time, the USCG pencil and paper multiple-choice items are studied extensively in academies, training schools, and by individuals on their own. In the best cases, this study may involve analysis of the Rules and have positive results on the understanding of their meaning and intent. In the worst cases, we have heard anecdotal evidence that some candidates memorize the available items with no real comprehension of the Rules and corresponding principles of action. Such memorization is largely divorced from reality at sea and requires no decision-making other than selection of a memorized choice.

We feel that an interactive test could have positive effects on training. Preparation for the interactive examination requires that the candidate examine a real scenario and make real-time decisions based on the situation presented. The difficulty in preparing a large data base of scenarios means that there is a possibility that candidates may attempt to "memorize" scenarios. In many respects this type of memorization is preferable to the present situation because the interactive examination offers the opportunity to memorize processes that will be useful at sea. In preparing the candidate must: 1.) memorize the proper application of the ROR, 2.) memorize a correct decision process in a real-time situation, and 3.) memorize the relationship of the ROR to operational actions such as keeping a proper lookout.

4. Preparation for test-taking. We heard a concern that potential candidates be given an opportunity for familiarization with the testing approach. We feel that this is not a difficult problem. If the tester were to be software that would run on a standard PC, a familiarization exercise on a diskette could be made available for potential candidates' use anywhere. (An opportunity for familiarization would be especially important for candidates who are unaccustomed to computer use. However, over time, as the amount of automation onboard vessels increases, the number of such candidates should decrease.)

5. Documentation of individual performance. A need to provide documentation of the actual test taken by an individual and his/her performance should be possible to provide. See Appendix A for recommendations for software capability.

7.0 SUMMARY OF RECOMMENDATIONS

We recommend further development of an interactive tester using a desk-top simulator for the assessment of ROR competence. Our recommended approach to this development is summarized in Section 5 above. The approach is sufficiently general to apply to the assessment of other mariner competencies, whether using desk-top simulators, part-task trainers, real equipment, or other settings. Specific recommendations for the development of the ROR tester are presented in Appendix A.

8.0 A RELATED STUDY OF ASSESSMENT OF COMPETENCE

The 1995 revision of the IMO STCW Code (International Maritime Organization, 1995) includes a new “specification of minimum standard of competence for officers in charge of a navigational watch” (and of an “engineering watch”). The Code presents a list of required competencies to be demonstrated by the officer, including a “Thorough knowledge of the content, application and intent of the International Regulations for Preventing Collisions at Sea.” We believe that the study reported here has made a beginning in the exploration of the issues involved in such a demonstration. Beyond the demonstration of ROR knowledge and application, our findings also provide a beginning to an understanding of the general issue of “demonstrating” competence.

The USCG Research and Development Center is presently engaged in another study for the National Maritime Center, Marine Examination Administration Branch (NMC-4B), a study entitled, “Qualifications and Training.” The purpose of the current phase of the study is to consider the STCW requirements for the officer in charge of a navigational watch and to determine how to demonstrate competence in a number of functions. We will: select a sample of the competencies listed in Table A-II-1 of the STCW Code; elaborate on the details of the “knowledge, understanding, and proficiency” that they require; examine and select appropriate methods for “demonstrating competence;” and develop “criteria for evaluating competence.”

The IRORT study, reported here, has suggested development approaches that will be considered further in the new study. Among these are approaches for: specifying comprehensive testing objectives, including test and scenario content; selecting or developing equipment that will support the testing objectives; and developing performance measures and standards. The combined study efforts should provide specific recommendations on assessment for a number of selected competencies for the officer in charge of a navigational watch and a general approach to developing assessment procedures for additional competencies required in the Code.

REFERENCES

Anastasi, A. (1988). Psychological Testing, sixth edition. New York: Macmillan Publishing.

Berk, R.A. (Ed.) (1984). A Guide to Criterion-Referenced Test Construction. Baltimore: The Johns Hopkins University Press.

Code of Federal Regulations. (1994). Title 46 CFR Shipping, Part 10-Licensing of Maritime Personnel. Revised as if October 1, 1994.

Gardenier, J. S., F.J. Flyntz, R.S. Spears, R.M. Willis, and D.C. North. (1987). A Desktop Simulator For Licensing Mariners. Proceedings of the Fourth International Conference on Marine Simulation (MARSIM-87). Trondheim, Norway, June 22-24, 1987.

Gatewood, R.D., and H.S. Feild. (1994). Human Resource Selection, third edition. Fort Worth, TX: The Dryden Press. (Harcourt Brace College Publishers).

Hammell, T.J., J.W. Gynther, J.A. Grasso, and M.E. Gaffney. (1981). Simulators for Mariner Training and Licensing Phase 2: Investigation of Simulator Characteristics for Training Senior Mariners. U.S. Coast Guard Report Number CG-D-08-82. CAORF Report Number 50-7915-02. NTIS Number AD A114746.

Hanley, M.J. (1984). Results of the Part-task Shiphandling Trainer Model Training Effectiveness Evaluation (TEE). Prepared by Ship Analytics, Incorporated, North Stonington, CT, 06359. Prepared for Naval Training Equipment Center, Human Factors Research Laboratory, (Code N-71), Orlando, FL, 32813.

Harris, R.J. (1975). A Primer of Multivariate Statistics. New York: Academy Press

Hughes, T. (1993). Officer of the Watch Simulator. Seaways, June 1993.

International Maritime Organization. (1995). International Convention on Standards of Training, Certification and Watchkeeping for Seafarers, 1978. Seafarers' Training, Certification and Watchkeeping (STCW) Code. STCW 6/Circ.1, 24 July 1995. (Table A-II/1.)

Jensen, O.A. and B. Shimberg. (1970). Developing Improved Examinations and Procedures for Licensing Deck and Engineering Officers in the U.S. Merchant Marine: Test Specifications. Prepared by Educational Testing Service, Princeton, NJ. Prepared for U.S. Coast Guard. September, 1970.

Kane, M.T. (1982). The validity of licensure examination. American Psychologist, 37, 911-918.

Kelly, M.J. (1988). Performance Measurement during Simulated Air-to-Air Combat. Human Factors, 1988, 30 (4), 495-506.

Mariner Licensing Device Final Report. (1987). Prepared by Engineering Computer Optecnomics, Inc., 1036 Cape St. Clair Center, Annapolis, MD 21401. For United States Coast Guard, Washington, DC 20593, 27 May 1987.

McCallum, M.C., M.W. Smith, G. Sandberg, D. Hard, R. Meurn, and R.D. Stewart. (1996). Criterion-based Assessment of Marine Simulator Operator Performance. Proceedings of the International Conference on Marine Simulation Towards Safer Seas and Cleaner Oceans. February 23-27, 1996, Alexandria, Egypt.

McCallum, M.C., M.W. Smith, G. Sandberg, D. Hard, and R.D. Stewart. (1995). Feasibility Assessment of an Interactive Rules of the Road Tester. Proceedings of International Symposium - Human Factors On Board. November 15-17, 1995, Bremen, Germany.

Meister, D. (1985). Behavioral Analysis and Measurement Methods. New York: John Wiley and Sons.

Meurn, R.J. (1995). Simulation Training at Our Nation’s Federal and State Maritime Academies. Proceedings of the Marine Safety Council -- November - December 1995. Pages 35 - 36.

Meurn, R.J. (1990). Watchstanding Guide for the Merchant Mariner. Centreville, Maryland: Cornell Maritime Press. Pages 152 - 172.

Moynehan, J.D., M.J. Hanley, and V. Pittsley. (1985). Results Of The Shiphandling Table-Top (Captains) Training Effectiveness Evaluation (TEE). (Performing Org. Rep No. 84-U-288) Prepared by Ship Analytics Inc., North Stonington CT. Prepared for the Naval Training Equipment Center, Orlando, FL, February, 1985.

National Research Council. (1996) Simulated Voyages. Washington, D.C.: National Academy Press.

Oil Pollution Act of 1990 (OPA 90). Public Law 101-380.

PC Maritime Ltd. (1993a). Officer Of The Watch Course Designer Manual. PC Maritime Ltd., Plymouth, UK. April 1993.

PC Maritime Ltd. (1993b). Officer Of The Watch Simulator Manual. PC Maritime Ltd., Plymouth, UK. May 1993.

Sandberg, D.R., R.D. Stewart, M.W. Smith, and M.C. McCallum. (1996). Development of an Interactive Assessment Device for Testing Rules of the Road Proficiency Using Part Task Simulation. Proceedings of Ninth International Navigation Simulator Lecturers' Conference. Vancouver, Canada. August 1996

Sandberg, G. and R.D. Stewart. (1995/6). Interactive Rules of the Road Testing. Seaways, December 1995 and January 1996.

Shimberg, B. (1981). Testing for Licensure and Certification. American Psychologist, 36, 1138-1146.

SPSS, Inc. (1990). SPSS/PC+ Advanced Statistics 4.0. Chicago: SPSS Inc.

Stewart, R.D., G. Sandberg, R. Meurn, and D. Hard. (1994). Interactive Rules of the Road Tester: Phase 1. Prepared by United States Merchant Marine Academy, Department of Marine Transportation, Kings Point, NY, 11024 for United States Coast Guard Research and Development Center, Groton, CT, 06340-6096.

Vreuls, D., and R.W. Obermayer. (1985). Human-system Performance Measurement in Training Simulators. Human Factors, 1985, 27(3), 241-25.

United States Coast Guard. (1996). An Exploration of the Feasibility and Benefit of Mariner Competency Testing Using a Desk-top Simulator: The Interactive Rules of the Road Tester Project. Prepared by the United States Coast Guard Research and Development Center, Groton, CT 06340-6096 for the United States Coast Guard, National Maritime Center, Arlington VA 22203-1804. February 1996.

United States Coast Guard. (1995). Navigation Rules. COMMANDANT INSTRUCTION M16672.2C. Washington, D.C.: U.S. Government Printing Office.

United States Coast Guard. (1993a). Licensing 2000 and Beyond. Office of Marine Safety, Security and Environmental Protection, U.S. Coast Guard, Washington DC 20593.

United States Coast Guard. (1993b). Rules of the Road Examination Module #05428-0200 dated February 1993.

[BLANK]

APPENDIX A

RECOMMENDED FEATURES FOR A DESK-TOP SIMULATOR

We identified the following system capabilities during the study as essential or desirable for testing the ability to apply ROR. We recommend their consideration and inclusion in future efforts to develop an interactive test of ROR. Some of these features have applicability for competence assessment beyond ROR and could profitably be included in computer-based systems for other training and testing applications. (The software that we used for our sample test did have many of the features listed here. See Section 2.2 for a brief description of its features. See also PC Maritime Limited, 1993a and 1993b; Hughes, 1993.)

- The software should have the capability to test for both Inland and International Collision Regulations.

- The software should provide substantial, user-friendly flexibility to the test designer. The designer should be able to develop scenarios with day, night, or fog conditions, various restricted waterway settings, and a variety of traffic ships. The designer's options should include denying the test taker's access to any system features when necessary for the test design. Other designer options are specified throughout this Appendix.

It would be desirable if the software produced a brief text summary of a scenario/test, listing the designer's specification of initial conditions, characteristics of own ship and traffic ships, scenario events, features denied the test taker, etc. This feature would assist in review of a test and would provide later documentation.

The designer should be able to insert written instructions and multiple-choice test questions, of considerable length, if necessary, into the scenarios. The designer should have the flexibility to specify such details as the length of the text, how much time should be allowed for answering, whether feedback should be given, etc. It would be desirable for the software to allow the designer to enter, review, edit, and print all the text in a scenario, without actually running the scenario.

- The software should automatically score multiple-choice items.

- The on-screen display should provide an out-the-window view with sufficient detail to allow the user to discern the aspect of traffic vessels and to identify the shapes and lights specified by the Rules. The finest detail could be provided by a "binocular" view. Background detail, such as water texture or clouds, is unnecessary in a ROR test and can be distracting on a small screen. There should be the capability for daytime, nighttime, and fog.

- There should be a variety of designer-selectable own ships. There should be the capability for multiple, varied traffic ships whose behavior can be specified by the test designer. Traffic ships should show the proper day shapes and lights and sound the proper sound signals for conditions. If traffic ships have the capacity to maneuver when in close quarters with own ship, the criteria for their maneuvers should be documented and available to the designer for consideration in the design of the scenarios. The designer should have the option to specify "rogue" ships that do not maneuver when in close quarters.

- The test taker should be provided with a variety of user-friendly capabilities, as they are allowed by the test designer. He/she should be able to change the direction of the out-the-window view to "look" in all directions and to use binoculars, bearing compass, and radar. He/she should be able to control the course and speed of own ship and to operate the sound signals. User-friendly software running on a standard PC probably means these features are represented by on-screen icons and controlled by keyboard and mouse. Scenario time should stop while instructions or questions are on the screen.

- The software should record all test taker actions in a log file and all the course and speed behavior of own ship and the traffic ships for later playback. Actions should be recorded with their time and duration; rotation of the view should be recorded with direction; and use of the bearing compass should be recorded with the bearing taken. The ship performance data recorded during a scenario should include, or be sufficient for later calculation of, the CPA to a threat vessel resulting from own ship's maneuver. Provision should also be made to determine the minimum CPA to all traffic ships in a scenario.

- If the ROR test is to be administered to more than one test taker at a time, there needs to be provision for shielding each screen from the other test takers and for headsets so that each test taker will hear sound signals only from his/her own simulator.

- During test administration, the software should allow the test taker to move quickly and easily between scenarios in a specified series. The system should automatically save performance on each scenario to the PC hard drive, without exiting the software.

- The software should allow the test designer considerable flexibility in the scoring of the scenarios. The software should support the development of a standard "catalogue" of "performance measures," representing the selection and processing of individual control actions over specified blocks of time. It should also allow the specification of scoring criteria and weights for these performance measures.

APPENDIX B:

TESTING OBJECTIVES FOR THE INTERACTIVE TEST

The first step in test design is generally the specification of testing objectives. For this study, the objectives were specified by extracting them from the USCG Rules of the Road Examination Third Mates Module # 05428-02200 (Stewart, Sandberg, Meurn, and Hard, 1994; United States Coast Guard, 1993b). Our Subject Matter Experts (SMEs) categorized the questions the Module contained and identified those suitable for incorporation into an interactive test. The questions that addressed only Inland Rules were discarded because the British-made software could not provide the appropriate conditions. Those categorized as factual/objective questions were presented to the cadets as multiple-choice questions on the computer screen. The SMEs categorized the remaining questions as recognition or operational and deemed them suitable for interactive testing. As many of these as possible were incorporated into three coherent, interactive scenarios, one daytime, one nighttime, and one fog. The analysis of the Examination Module is described briefly in Section 5 in the present report and in greater detail elsewhere (Sandberg and Stewart, 1995/6; Stewart, Sandberg, Meurn, and Hard, 1994).

The questions selected for incorporation in each scenario were matched to the International Rules that applied, as in the listing below. The phrases in bold type are those that we interpreted as requiring an operational or interactive demonstration of the concept. Only key words from the Rules are presented here. The full text of the Rules is available in COMDTINST M16672.2C (United States Coast Guard, 1995).

1. Daytime scenario:

Rule 7: Risk of Collision
(a) Every vessel shall **use all available means** appropriate to the prevailing circumstances and conditions to determine if risk of collision exists...
(b) **Proper use shall be made of radar equipment**...

Rule 8: Action to Avoid Collision
(a) Any action taken to avoid collision shall...**be positive, made in ample time** and with due regard to the observation of good seamanship.
(c) ...alteration of course alone may be the most effective action...provided that it is **made in good time, is substantial and does not result in another close-quarters situation**.

Rule 24: Towing and Pushing
(e) A vessel ...being towed shall exhibit:...when the length of the tow exceeds 200 meters, a diamond shape...

(g) An inconspicuous, partly submerged vessel or object...being towed, shall exhibit:...a diamond shape...

Rule 34: Maneuvering and Warning Signals
(a) ...a power-driven vessel underway, **when maneuvering** as authorized or required by these Rules, **shall indicate that maneuver by the following signals on her whistle**...

2. Nighttime scenario:

Rule 7: Risk of Collision
(a) Every vessel shall **use all available means** appropriate to the prevailing circumstances and conditions to determine if risk of collision exists...
(b) **Proper use shall be made of radar equipment**...

Rule 8: Action to Avoid Collision
(a) Any action taken to avoid collision shall...**be positive, made in ample time** and with due regard to the observation of good seamanship.
(c) ...alteration of course alone may be the most effective action... provided that it is **made in good time, is substantial and does not result in another close-quarters situation**.

Rule 24: Towing and Pushing
(a) A power-driven vessel when towing shall exhibit:...two masthead lights in a vertical line...

Rule 25: Sailing Vessels Underway and Vessels Under Oars
(a) A sailing vessel underway shall exhibit...

Rule 26: Fishing Vessels
(a) A vessel engaged in fishing...shall exhibit...

Rule 27: Vessels Not Under Command or Restricted in Their Ability to Maneuver
(a) A vessel not under command shall exhibit...
(b) A vessel restricted in her ability to maneuver...shall exhibit...

Rule 34: Maneuvering and Warning Signals
(a) ...a power-driven vessel underway, **when maneuvering** as authorized or required by these Rules, **shall indicate that maneuver by the following signals on her whistle**...

3. Fog scenario:

Rule 7: Risk of Collision
(a) Every vessel shall **use all available means** appropriate to the prevailing circumstances and conditions to determine if risk of collision exists...
(b) **Proper use shall be made of radar equipment**...

Rule 8: Action to Avoid Collision
(a) Any action taken to avoid collision shall...**be positive, made in ample time** and with due regard to the observation of good seamanship.

Rule 19: Conduct of Vessels in Restricted Visibility.
(e) ...every vessel which hears apparently forward of the beam the fog signal of another vessel...**shall reduce her speed**

Rule 35: Sound Signals in Restricted Visibility
(e) A vessel towed...if manned, shall...sound...one prolonged followed by three short blasts...

The specified test objectives guided the design of the interactive scenarios. Each included the traffic ships that would require the test taker to recognize the selected shapes, lights, or signals and each included an encounter that would require the test taker to demonstrate the responsibilities indicated in bold type. The content of each scenario is documented in greater detail in Appendix C which follows.

The test object objective also guided the selection of performance measures for each scenario. Multiple-choice questions were embedded in the scenarios to query the test taker on each required recognition of traffic vessels. Performance measures were developed for the interactive operations. The recognition questions and the performance measures for each scenario are documented in Appendix C.

[BLANK]

APPENDIX C

OPERATIONAL MEASURES AND PERFORMANCE STANDARDS

OVERVIEW OF THIS APPENDIX.

Our approach to the development of the operational measures and performance standards is described in the main test in Section 3.0. This appendix documents the USMMA subject matter experts' (SME) specification of measures and performance standards for those measures, for each of the interactive scenarios -- daytime, nighttime, and fog.

THE OPERATIONAL MEASURES.

For each scenario, a summary of the initial conditions for own ship and the traffic vessels, major events, and embedded recognition items is provided here. For each scenario, the basic testing objectives were the same, that the candidate demonstrate an understanding of the three basic requirements of navigation law during watch standing. These requirements are to:

1. Maintain a good lookout/determine if risk of collision exists
2. Take action or maneuver to avoid collision
3. Determine if own ship's action or maneuver was adequate to avoid collision/determine that action or maneuver does not put own ship in a close quarters situation with other vessels

The specific operational measures that represent understanding of these requirements are somewhat different for each scenario.

THE PERFORMANCE STANDARDS.

In setting the performance standards, two different scales were used. In most cases, a three-point scale of *Proficiency Level* was applied. In selected cases, a two-point scale of *Competency Level* was applied.

Proficiency Level is defined with reference to both navigation law and professional standards. Considering both, it is the consistency of performance with legally mandated actions, as defined by navigation law; and with the indicated level of prudent seamanship (expert, qualified, unqualified) in the operational application of navigation law. These three proficiency levels are:

Expert: Performance is fully consistent with all legal mandates and meets the highest professional standards of prudent seamanship in the operational application of navigation law

Qualified:	Performance is fully consistent with all legal mandates and meets acceptable professional requirements of prudent seamanship in the operational application of navigation law
Unqualified:	Does not meet one or both of the legally mandated actions and/or acceptable professional requirements of prudent seamanship in the operational application of navigation law

Competency Level is defined with reference to navigation law. It is the consistency of performance with legally mandated actions, as defined by navigation law. The two levels of competency are:

Competent:	Performance is fully consistent with legally mandated actions
Incompetent:	Performance is inconsistent with legally mandated actions

DAYTIME SCENARIO

Initial message is as follows:

It is 0900 and you are on watch aboard a tanker at full sea speed (14.9 knots) steering 000 true on international waters. You are not in or near an area of restricted visibility. You have four vessels in sight. Identify each of the vessels, using the binoculars if necessary. You have no control of your vessel's movements (course and speed) at this time. You may be asked questions concerning any of these vessels

.

Initial situation of other vessels is as follows:

Vessel	**Bearing (˚ R)**	**Distance (nm)**	**Course (˚ T)**	**Speed (EOT)**	**Constraints**
1. Tug & tow	340	1.6	000	Full ahead	tow astern tow submerged tow > 200m
2. Fishing vessel	354	2.7	270	dead slow ahead	fishing gear on port side
3. Container vessel (threat)	105	4.0	333	full ahead	None rogue vessel
4. VLCC	352	8.2	180	Full ahead	constrained by draught

The sequence of messages and questions that follow is:

09:04:54 *The vessel broad on your port bow is a ______.*
c. tug towing a partly submerged object, with the length of the tow exceeding 200 m. (1.)

09:05:17 *The vessel two points on your port bow is a _____.*
b. vessel engaged in fishing, with outlying gear extending more than 150 m. (2.)

09:05:40 *With what vessel(s) is/are there risk of collision? You may select more than one answer.*
d. vessel 1 point abaft your starboard beam (3.)

09:06:10 *You now have full control of your vessel. (engine and rudder) You may take whatever action you deem necessary.*

(At this point, own ship could maneuver to increase the projected CPA with the container vessel that has continued to close at a bearing of 105 ˚ R and is at a range of ~3 nm).

DAYTIME SCENARIO PERFORMANCE MEASURES

1. Maintain a good lookout/determine if risk of collision exists

1.1 Percentage of total visual search time allocated to each of the four quadrants during the first six minutes of the test scenario.
Comments: Six minutes for a radar plot is the only standard "out there." It is a fundamental tenth of an hour. Here, it is applied to visual search. Note that the percentage of time must fall within the envelope in all four quadrants to achieve the performance level specified. (Using OOW's log, visual time in each direction should be calculated for first 6 minutes and radar time subtracted from that amount to obtain the desired value.)

Percentage Range of Total Visual Search Time in Each Direction During First Six Minutes				**Performance Level**
Forward	**Starboard**	**Aft**	**Port**	
35-50	*30-45*	*5-15*	*5-15*	**Expert**
25-80	*15-80*	*5-35*	*5-35*	**Qualified**
Neither Expert nor Qualified				**Unqualified**

1.2 Binocular viewings of vessels to be completed during the first six minutes of the test scenario.

Number of Vessels Viewed	**Number of Times Each Vessel is Viewed**	**Range of Time to Complete Viewing of Vessels**	**Performance Level**
4	*2*	*6 min.*	**Expert**
4	*1*	*6 min.*	**Qualified**
Neither Expert nor Qualified			**Unqualified**

1.3 Visual bearings of vessels to be completed during the first six minutes of the test scenario.

Number of Bearings on Diff. Vessel	**Number of Bearings on Each Vessel**	**Range of Time to Complete Bearings**	**Performance Level**
4	*3*	*6 min.*	**Expert**
4	*2*	*6 min.*	**Qualified**
Neither Expert nor Qualified			**Unqualified**

1.4 Percent of time **and** number of times radar viewed during the first six minutes of the test scenario. (Note that both conditions must be met to obtain an Expert or Qualified score.)

Percent of Time Spent Radar Viewed	**Number of times radar viewed**	**Performance Level**
20-40	*2 or more*	**Expert**
15-20 or 40-50	*2 or more*	**Qualified**
Neither Expert nor Qualified		**Unqualified**

2. Take action or maneuver to avoid collision

2.1 Action/maneuver resulting in an adequate CPA with Threat Vessel following completion of maneuver. (Comment: Although this measure was considered extremely important by the SMEs, it was not possible to extract this information from the record in order to include it in the analysis.)

Action/Maneuver			**Resulting CPA**	**Performance Level**
Left Turn	**Right Turn**	**Slow or Reverse**		
No	*Yes, in 8 min. or less*	*Yes*	*1.5 nm or more*	**Expert**
No	*Yes, in 8-13 min.*	*Yes*	*0.5 - 1.5 nm*	**Qualified**
Neither Expert nor Qualified				**Unqualified**

2.2 Timely sounding of correct signal at time of maneuver.

Action	**Competency Level**
Correct signal sounded within 30 seconds before or after maneuver	**Competent**
No signal or incorrect signal	**Incompetent**

3. Determine if own ship's action or maneuver was adequate to avoid collision/determine that action or maneuver does not put own ship in a close quarters situation with other vessels

3.1 Avoid close quarter situation with all other vessels at all times following action or maneuver.

Minimum CPA During Entire Scenario	Performance Level
More than 0.5 nm	**Expert**
0.25 - 0.5 nm	**Qualified**
Less than 0.25 nm	**Unqualified**

3.2 Visual bearing of vessels after maneuver completion.

Bearings on Number Diff. Vessels	Number of Bearings on Each Vessel	Range of Time to Complete Bearings	Performance Level
4	*2*	*Within 4 minutes after slowing or course change commences*	**Expert**
Container vessel (threat)	*2*	*Within 8 minutes after slowing or course change commences*	**Qualified**
Neither Expert nor Qualified			**Unqualified**

3.3 Percent of time **and** number of times radar viewed during the first six minutes following the maneuver. (Note that both conditions must be met to obtain an Expert or Qualified score.)

Percent of Time Spent Radar Viewed	Number of times radar viewed	Performance Level
20-40	*2 or more*	**Expert**
15-20 or 40-45	*2 or more*	**Qualified**
Neither Expert nor Qualified		**Unqualified**

NIGHTTIME SCENARIO

Initial message is as follows:

You are on watch aboard a container vessel at full sea speed (19.5 knots), steering a course of 000 true on international waters. Visibility is not restricted. You have six vessels in sight. Identify each of these vessels, using the binoculars if necessary. You have no control of the vessel's movements (course or speed) at this time. You will be asked questions concerning each of the six vessels

Initial situation of other vessels is as follows:

Vessel	Bearing (˚ R)	Distance (nm)	Course (˚ T)	Speed (EOT)	Constraints
1. Tug & tow	350	3.0	180	slow ahead	length of tow > 200m
2. Trawler	015	2.0	000	dead slow ahead	fishing and shooting nets
3. Tug & tow	210	0.5	010	full ahead	tow alongside on starboard side
4. Tanker	340	2.0	170	dead slow ahead	not under command
5. Yacht	280	6.0	000	slow ahead	under sail
6. Coaster	345	8.0	189	slow ahead	none rogue vessel
7. VLCC (threat)	048	6.1	300	full ahead	restricted in ability to maneuver rogue vessel

The sequence of messages and questions is as follows:

04:00:00 Scenario starts at 04:00:00

04:03:18 *What is the vessel on your port beam?*
d. sailboat underway (5.)

04:03:48 *The vessel fine on your port bow is* _____
c. vessel towing, length of tow exceeding 200 meters. (1.)

04:04:18 *The vessel 3 points on your starboard bow is* ____
d. vessel engaged in trawling, shooting nets (2.)

04:05:03 *The vessel broad on your starboard bow is* ____
a. vessel restricted in ability to maneuver, making way (7.)

04:05:48 *The vessel dead astern is* ____
d. vessel towing alongside, the barge fast to tug's starboard side (3.)

04:06:33 *The vessel 3 points abaft your beam is* ____
a. a vessel not under command (4.)

04:07:18 *Is there a risk of collision with any of the other vessels? If so, which one(s)?*
c. the large vessel broad on your stb. bow (7.)

04:07:33 *You now have full control of your vessel*

The VLCC is approaching from starboard. If own ship does not maneuver, at 04:21:00 the CPA of threat vessel will be 0.11 nm.

NIGHTTIME SCENARIO PERFORMANCE MEASURES

1. Maintain a good lookout/ determine if risk of collision exists

1.1 Percentage of total visual search time allocated to each of the four quadrants during the first 7.5 minutes of the test scenario:
Comments: Six minutes for a radar plot is the only standard "out there." It is applied to visual search for Daytime scenario. For the Nighttime scenario, more time is allowed because there are more target ships. Note that the percentage of time must fall within the envelope in all four quadrants to achieve the performance level specified. (Using OOW's log, visual time in each direction should be calculated for first 7.5 minutes and radar time should be subtracted from that amount to obtain the correct value.)

Percentage Range of Total Visual Search Time in Each Direction During First 7.5 Minutes				**Performance Level**
Forward	**Starboard**	**Aft**	**Port**	
35-50	*30-45*	*5-15*	*5-15*	**Expert**
25-80	*15-80*	*5-35*	*5-35*	**Qualified**
Neither Expert nor Qualified				**Unqualified**

1.2 Binocular viewings of vessels to be completed within the following period of time from the start of the scenario:

Number of Vessels Viewed	**Number of Times Each Vessel Viewed**	**Time to Complete Viewing of Vessels**	**Performance Level**
7	*2*	*7.5 min.*	**Expert**
7	*1*	*7.5 min.*	**Qualified**
Neither Expert nor Qualified			**Unqualified**

1.3 Visual bearings of vessels to be completed during the first 7.5 minutes of the test scenario:

Bearings on Number Diff. Vessels	Number of Bearings on Each Vessel	Range of Time to Complete Bearings	Performance Level
7	*3*	*7.5 min.*	**Expert**
7	*2*	*7.5 min.*	**Qualified**
Neither Expert nor Qualified			**Unqualified**

1.4 Percent of time **and** number of times radar viewed during the first 7.5 minutes of the test scenario: (Note that both conditions must be met to obtain an Expert or Qualified score.)

Percent of Time Spent Radar Viewed	Number of times Radar Viewed	Performance Level
20-40	*2 or more*	**Expert**
15-20 or 40-50	*2 or more*	**Qualified**
Neither Expert nor Qualified		**Unqualified**

2. Take action or maneuver to avoid collision

2.1 Action/maneuver resulting in an adequate CPA with Threat Vessel following completion of maneuver:

Action/Maneuver				
Left Turn	**Right Turn**	**Slow or Reverse**	**Resulting CPA**	**Performance Level**
No	*Yes, in 9.5 min. or less*	*Yes*	*1.5 nm or more*	**Expert**
No	*Yes, in 9.5 -15 min.*	*Yes*	*0.5 - 1.5 nm*	**Qualified**
Neither Expert nor Qualified				**Unqualified**

2.2 Timely sounding of correct signal at time of maneuver.

Action	Competency Level
Correct signal sounded within 30 seconds	**Competent**
No signal or incorrect signal	**Incompetent**

3. Determine if own ship's action or maneuver was adequate to avoid collision/determine that action or maneuver does not put own ship in a close quarters situation with other vessels

3.1 Avoid close quarter situation with all other vessels at all times following action or maneuver.

Minimum CPA During Entire Scenario	**Performance Level**
More than 0.5 nm	**Expert**
0.25 - 0.5 nm	**Qualified**
Less than 0.25 nm	**Unqualified**

3.2 Visual bearing of vessels after maneuver completion.

Bearings on Number Diff. Vessels	**Number of Bearings on Each Vessel**	**Range of Time to Complete Bearings**	**Performance Level**
7	*2*	*Within 5 minutes after slowing or course change commences*	**Expert**
VLCC (threat)	*2*	*Within 9 minutes after slowing or course change commences*	**Qualified**
Neither Expert nor Qualified			**Unqualified**

3.3 Percent of time **and** number of times radar viewed during the first six minutes following the maneuver. (Note that both conditions must be met to obtain an Expert or Qualified score.)

Percent of Time Spent Radar Viewed	**Number of times Radar Viewed**	**Performance Level**
20-40	*2 or more*	**Expert**
15-20 or 40-45	*2 or more*	**Qualified**
Neither Expert nor Qualified		**Unqualified**

FOG SCENARIO

Initial message is as follows:

It is 1000 and you are on watch aboard a tanker on international waters. Visibility is reduced to less than 0.5 miles by fog. You must sound the appropriate fog signals. Your course is 000 degrees and your speed is 11.2 knots (half ahead). Use all means available to determine the presence of other vessels. You have no control of your vessel's course and speed ***at this time****.*

Initial situation of other vessels is as follows:

Vessel	**Bearing (˚ R)**	**Distance (nm)**	**Course (˚ T)**	**Speed**	**Constraints**
1. Tug & tow	135	0.25	000	full ahead	towing astern tow length > 200m
2. VLCC	333	3.0	270	stop	no constraint
3. Container (threat) (materializes at 10:12:00)	000	2.0	180	dead slow ahead	no constraints rogue vessel

The sequence of events and messages is as follows:

10:00:00 *Part IV - click the OFF/On icon with the mouse to start the examination.*

10:00:05 *Instruction #1* (see initial conditions above)

10:00:07 *Instruction #2 Sound Signals*

You can sound all the prescribed whistle signals for fog and clear weather: short blast are sounded by pressing the appropriate key. ([1] for one short blast, [2] for two short blast, and so on) Prolonged blast are sounded by holding [Alt] and pressing the appropriate number key. ([Alt 1] for one prolonged blast, [Alt 2] for two prolonged blast, and so on) To make special signals press any keyboard letter to sound the equivalent Morse sound signal. (for example, pressing U will sound ..-, pressing D will sound -.., and so on).

10:01:00+ sound: one prolonged blast, two short, pause, one prolonged, three short

10:02:00+ sound: one prolonged blast, two short, pause, one prolonged, three short

10:02:24 Tug & tow (1.) changes course to 000 ˚ True, dead slow ahead

10:05:00+ sound: one prolonged blast, two short, pause, one prolonged, three short

10:06:00 *The lookout reports hearing a fog signal on the starboard quarter.*

10:06:27 *The vessel on your starboard quarter is sounding signal indicating it is a ______.*
b. a tug towing a manned vessel. (1.)

10:08:00 sound: two prolonged blasts

10:08:00 sound: two prolonged blasts

10:12:00 *The lookout reports hearing fog signals on the port bow.*

10:12:00 VLCC (2.) vanishes/dematerializes

10:12:05 *You hear a sound signal of a vessel on your port bow. What type of vessel could it be?*
d. a power driven vessel underway, stopped, making no way through the water. (2.)

10:12:30 *You now have full control of the course and speed of your vessel and may take any action you deem necessary.*

10:13:00 *Radar Failure, Press any key.*

10:14:00 sound: one prolonged blast

10:14:59 *The lookout reports hearing a fog signal forward of the beam.*

10:16:48 sound: one prolonged blast

10:19:00 sound: one prolonged blast

10:19:30 Rogue container vessel (3.) appears on a reciprocal course of 225 ˚ True, dead slow ahead If own ship did not slow immediately after radar failure, there will be a collision .

THE END OF PART IV
Stop the simulation by clicking the "ON/OFF" icon with the mouse. CALL THE EXAM PROCTOR.

FOG SCENARIO PERFORMANCE MEASURES

1. Maintain a good lookout/determine if risk of collision exists

1.1 Percentage of total visual search time allocated to each of the four quadrants during the first 12 minutes of the test scenario.
(Comments: Six minutes for a radar plot is the only standard "out there." It is applied to visual search in the Daytime scenario. More time is allowed here to allow the test taker to hear the fog signals of each of the traffic vessels twice. Note that the percentage of time must fall within the envelope in all four quadrants to achieve the performance level specified. (Using OOW's log, visual time in each direction should be calculated for first 12 minutes and radar time should be subtracted from that amount. Note also that at start of scenario, radar shows one vessel at 0.25 nm on starboard quarter. However, OOW makes it difficult to look in this direction. Forward of the beam requires both forward and beam. This need was considered when the following percentages were selected.)

Percentage Range of Total Visual Search Time in Each Direction During First 12 Minutes				Performance Level
Forward	**Starboard**	**Aft**	**Port**	
20-30	*20-30*	*20-30*	*20-30*	**Expert**
20-40	*10-30*	*10-30*	*20-40*	**Qualified**
Neither Expert nor Qualified				**Unqualified**

1.2 Percent of time **and** number of times radar viewed during the first 12 minutes of the test scenario. (Note that both conditions must be met to obtain an Expert or Qualified score.)

Percent of Time Spent Radar View	Number of Times Radar Viewed	Performance Level
20-40	*6 or more*	**Expert**
40-60	*3 or more*	**Qualified**
Neither Expert nor Qualified		**Unqualified**

1.3 Visual search to **starboard** and **aft** following lookout report of fog whistle off starboard quarter.

Maximum Elapsed Time	**Performance Level**
30 sec	**Expert**
40-60 sec	**Qualified**
doesn't look	**Unqualified**

1.4 Visual search to **port** and **forward** following lookout report of fog whistle off forward beam.

Maximum Elapsed Time	**Performance Level**
30 sec	**Expert**
40-60 sec	**Qualified**
doesn't look	**Unqualified**

1.5 Restricted visibility sound signal usage

.

Signal Usage	**Performance Level**
Sounding of correct signal within 1 minutes of scenario start	**Expert**
Sounding of correct signal within 1-3 minutes of scenario start	**Qualified**
Sounding of none/incorrect signal	**Unqualified**

2. Take action or maneuver to avoid collision

2.1 Substantial and timely reduction in speed following radar failure at 10:13.

Speed Reduced To:	Time After Radar Failure	Performance Level
Stop or astern	*30 sec*	**Expert**
Dead slow	*30 sec*	**Qualified**
Not reduced	*Not reduced*	**Unqualified**

2.2 Predominantly looking ahead from 10:14:59 when lookout reports hearing a fog signal forward of the beam until vessel comes into view. (Note that time vessel comes into view is dependent on own vessel speed. Note also that in OOW looking forward of the beam requires three directions.)

Percent Time Spent Looking Ahead	Performance Level
80-90	**Expert**
70-80 or 90-95	**Qualified**
<70 - >95	**Unqualified**

2.3 Closest CPA with any vessel during scenario

CPA Range	Performance Level
No collision	**Competent**
Collision	**Incompetent**