

A METHOD FOR DEVELOPING MARINER ASSESSMENTS

Chapter 3

A MANUAL FOR ASSESSMENT DEVELOPERS

This chapter describes a systematic, step-by-step method for developing Performance-based Assessment (PBA) of mariner proficiencies.

This report may be downloaded from the U.S. Coast Guard Research and Development Center web site at <http://www.rdc.uscg.mil>.

ACKNOWLEDGMENTS

This manual represents the current point in an ongoing research and development project conducted by the United States Coast Guard (USCG) Research and Development Center for the USCG National Maritime Center over several years. The following sponsors, academy representatives, and industry representatives, and colleagues supported the development of this manual.

Mr. Perry Stutman of the United States Coast Guard National Maritime Center has provided continual support and guidance in the conduct of this project. Other members of the National Maritime Center, including Captain William Bennett, Mr. John Bobb, Mr. David Field, and Mr. Albert Kirchner also provided support and guidance during the course of this work.

Participants of early work that significantly influenced the orientation of the current effort include Captains George Sandberg, Richard Stewart, Douglas Hard, and Robert Meurn, all faculty members of the United States Merchant Marine Academy. More recently, members of maritime academies have collaborated on the development of sample assessment procedures which influenced the content of this manual. Significant contributions have been made by Captain Joseph Murphy and Ms. Emily Sporn of Massachusetts Maritime Academy; and Dr. Paul Jackson and Mr. Robert Hammaker of California Maritime Academy.

Representatives from the maritime industry have also reviewed early version of this manual. Special thanks go to Captains Thomas Blanchard and Nils Knutstad of SeaRiver Maritime, Inc.

CHAPTER 3 – A MANUAL FOR ASSESSMENT DEVELOPERS

Introduction.....	3-5
Instructions	3-6
1.0 Specify Assessment Objectives	3-7
1.1 Identify regulatory requirements	3-7
1.2 Analyze job requirements	3-7
1.3 Describe individual assessment objectives	3-7
2.0 Determine Assessment Methods	3-8
2.1 Identify alternatives assessment methods	3-8
2.2 Review the advantages and disadvantages of alternative assessment methods	3-9
2.3 Determine assessment methods	3-10
3.0 Specify Assessment Conditions	3-11
3.1 Describe assessment settings or scenarios	3-11
3.2 Specify oral and written questions.....	3-12
4.0 Develop Proficiency Criteria	3-13
4.1 Identify components of assessment objectives	3-13
4.2 Develop individual measures and standards	3-14
4.3 Develop proficiency criteria and scoring procedures	3-19
4.4 Validate measures, standards, proficiency criteria, and scoring procedures	3-22
5.0 Prepare the Assessment Materials	3-22
5.1 Prepare detailed assessment worksheets.....	3-23
5.2 Prepare assessor instructions	3-23
5.3 Prepare candidate instructions	3-24
5.4 Finalize oral and written questions	3-24
5.5 Finalize simulator and/or shipboard procedures.....	3-25
5.6 Document references, source materials, and the validation process	3-25
Implementation.....	3-25
Glossary	3-27
References.....	3-29

LIST OF ILLUSTRATIONS

Figure 1. General process for developing measures and standards	3-16
Figure 2. General process for developing proficiency criteria and scoring procedures	3-20

LIST OF TABLES

Table 1. Summary of <i>Lookout</i> assessment objectives	3-9
Table 2. General guidance on the selection of assessment methods	3-10
Table 3. Assessment method advantages and disadvantages	3-11
Table 4. <i>Lookout</i> assessment objectives and methods.....	3-12
Table 5. Summary of assessment conditions for <i>Lookout</i> assessment Objective 4.....	3-13
Table 6. Oral questions from the <i>Lookout</i> assessment	3-14
Table 7. Recommended information sources for identifying components of assessment objectives	3-15
Table 8. Examples of three categories of mariner actions or outcomes	3-17
Table 9. Example of a single measure and standard applied to an assessment action	3-19
Table 10. Example of multiple standards applied to a single action	3-20
Table 11. Examples of hypothetical assessment scores and results	3-22

INTRODUCTION

Meeting the requirements of the *Seafarers' Training, Certification and Watchkeeping Code (STCW Code)* is a current challenge of the United States maritime industry. The *STCW Code* directs the United States maritime industry to develop procedures for assessing mariner proficiency on the basis of practical demonstration. The *STCW Code* defines areas for which mariners must receive certification and specifies the need to assess proficiency on the basis of skill and knowledge demonstrations by the mariner. However, there has been an ongoing debate regarding what, precisely, is implied by the *STCW Code* directive for assessments that address:

. . . the proper performance of functions onboard ship in accordance with the internationally agreed criteria as set forth herein and incorporating prescribed standards or levels of knowledge, understanding, and demonstrated skill . . . (*STCW Code*, Section A-1/1, paragraph 1.1).

This chapter is intended to be used as a manual and be considered as such and is referred to as such throughout this chapter. The intent of this manual is to outline a process that can be used by professional mariners in training institutions and other organizations to develop mariner assessment procedures. The process described in this manual conforms to international standards and domestic regulations, especially the International Maritime Organization's *STCW Code* and the U.S. Coast Guard's *Navigation and Vessel Inspection Circulars (NVICs)* that address implementation of the *STCW Code* within the United States. The reference section of this manual lists specific *STCW* documents and applicable *NVICs*.

This manual is based on a comprehensive methodology for mariner assessment development documented in the first chapter of this report. The methodology is also presented in a set of workshop materials and exercises entitled *A Workshop for Assessment Developers* (Chapter 2). More generally, this manual relies heavily on the general method of *Instructional Systems Development (ISD)*. More detailed reference information on all of the above sources is provided in the reference section of this manual.

The objective behind the process outlined in this manual is to develop *valid* and *reliable* mariner assessment procedures. Valid assessments evaluate mariners on the basis of their demonstration of job-critical knowledge, skills, and abilities. Valid assessments serve as accurate predictors of a mariner's abilities to fulfill the functional requirements of a job or task. Reliable assessments consistently obtain the same assessment results when different mariners with comparable skills are assessed. An assessment must be reliable and evaluate job-critical requirements in order to be valid.

The format of an assessment procedure is not essential to the validity and reliability of that procedure. Differences in organizational goals, assessment objectives, and performance standards preclude prescribing a single format for assessments. However, a sample assessment that proved to be useful during the development of this process is provided in this manual as an aid to the reader.

The remainder of this manual consists of a set of instructions for developing mariner assessment procedures, a glossary, a reference section, and a sample assessment procedure that is referred to

throughout the instructions. The instructions are intended to provide a relatively straightforward presentation of the steps needed to develop a valid and reliable assessment procedure. Issues critical to the validity and reliability of an assessment procedure are discussed, as necessary. The glossary provides a set of terms and definitions that are internally consistent with one another and, to the extent possible, consistent with current STCW, U.S. Code of Federal Regulations (CFR), and NVIC terms and definitions. The reference section identifies the international and domestic regulatory source documents; other manuals, authoritative books, and articles that address assessment development; and selected recent research on mariner assessment sponsored by the U.S. Coast Guard (USCG).

Instructions

This manual identifies a series of steps that can be conducted to develop a valid and reliable assessment procedure that meets International Maritime Organization (IMO) and USCG requirements for mariner demonstration of proficiency in selected areas. Following is an outline of the five basic steps and associated sub-steps described in this manual.

- 1.0 Specify Assessment Objectives.
 - 1.1 Identify regulatory requirements.
 - 1.2 Analyze job requirements.
 - 1.3 Describe individual assessment objectives.
- 2.0 Determine Assessment Methods.
 - 2.1 Identify alternative assessment methods.
 - 2.2 Review the advantages and disadvantages of alternative assessment methods.
 - 2.3 Determine assessment methods.
- 3.0 Specify Assessment Conditions.
 - 3.1 Describe assessment settings or scenarios.
 - 3.2 Specify oral and written questions.
- 4.0 Develop Proficiency Criteria.
 - 4.1 Identify components of assessment objectives.
 - 4.2 Develop individual measures and standards.
 - 4.3 Develop proficiency criteria and scoring procedures.
 - 4.4 Validate measures, standards, proficiency criteria, and scoring procedures.

5.0 Prepare the Assessment Materials.

5.1 Prepare detailed assessment worksheets.

5.2 Prepare assessor instructions.

5.3 Prepare candidate instructions

5.4 Finalize oral and written questions.

5.5 Finalize simulator and/or shipboard procedures.

5.6 Document references, source materials, and the validation process.

The following instructions are presented in a standard format to facilitate review and reference. Following the title of each step, the goal of the step is briefly described, followed by a listing of the sub-steps for that step. Then, separate instructions, discussions, and examples are presented to address the associated procedures, issues, and output.

1.0 Specify Assessment Objectives

Goal of this step

Identify the knowledge, skills, and abilities on which the mariner will be assessed.

Sub-steps

1.1 Identify regulatory requirements.

1.2 Analyze job requirements.

1.3 Describe individual assessment objectives.

1.1 Identify regulatory requirements

A mariner assessment procedure should be explicitly associated with a specific requirement in the *STCW Code*. The format of the *STCW Code* provides five levels of requirements that are readily identifiable, as shown in the following example:

Chapter: II – Standards regarding the master and deck department.

Table: A-II/1 – Specification of minimum standard of competence for officers in charge of a navigational watch on ships of 500 gross tonnage or more.

Function: Navigation at the operational level.

Competence: Plan and conduct a passage and determine position.

Knowledge, understanding and proficiency: Celestial navigation – Ability to use celestial bodies to determine the ship's position.

To specify the assessment objectives for an assessment procedure, begin by identifying all of the *STCW Code* requirements corresponding to these five levels. Later, this information will be included in the assessment documentation. In the *Lookout* assessment example in Appendix A, the *STCW Reference Information* entries in the *Assessment Control Sheet* provide similar

information from the *STCW Code*.

In developing an assessment to meet certification or licensing, other regulatory requirements should also be addressed. For example, the CFR provides explicit guidance regarding lifeboatman examination and demonstration of ability. If the assessment is intended to satisfy these additional regulatory requirements, the requirements should be explicitly addressed by the assessment and referred to in the assessment documentation.

1.2 Analyze job requirements

STCW Code and regulatory requirements typically will **not** provide sufficiently detailed information to support the development of a mariner assessment. To fill this gap, assessment developers should refer to technical manuals, job instructions, textbooks, and task analyses to aid in the identification of job requirements. If such sources are used to derive assessment objectives, they should be identified in the assessment documentation. For the present *Lookout* assessment example, the *STCW Code* specifies the following knowledge, understanding, and proficiency: “*Responsibilities of a lookout, including reporting the appropriate bearing of a sound signal, light, or other object in degrees or points.*” This information from the *STCW Code* was used to define the general requirements, which were augmented by a review of 46 CFR, Chapter 1, Subpart 97.27 – Lookouts, and by the *Navigation Rules* (Rule 2 – Responsibility and Rule 5 – Lookout). Rule 5 states: “*Every vessel shall at all times maintain a proper look-out by sight and hearing as well as by all available means appropriate in the prevailing circumstances and conditions so as to make a full appraisal of the situation and of the risk of collision.*”

In most cases, an individual assessment area will be analyzed to identify a set of several related assessment objectives that will be addressed separately during a proficiency assessment. Each of these objectives should represent a critical component of the job requirement being analyzed. At this point, the goal is to identify those broad areas of knowledge, skill, and ability that are critical to successful job performance. Later, during Step 4.0, a more detailed task analysis will be conducted. There are numerous job analysis methods that might be used in defining assessment objectives. Further guidance on the conduct of job and task analysis can be found in selected references listed in this manual (e.g., Jonassen, Hannum, & Tessmer, 1989; Mager, 1997; Rothwell & Kazanas, 1998). Review of available texts, along with review and discussion from qualified professionals, should result in an adequate set of assessment objectives.

An individual assessment procedure must meet the stated *STCW Code* requirements and be appropriate in scope in order to be practically feasible. An assessment should address a single job requirement or integrated set of job requirements defined in the *STCW Code*. In addition, the time and resource requirements for assessment should be within a practical level. An assessment procedure is too broad in scope if it addresses a wide range of unrelated job requirements defined in the *STCW Code* and/or requires an unnecessarily extensive amount of time and resources.

1.3 Describe individual assessment objectives

Each assessment objective should include a general description of the mariner performance that will be required during assessment and the conditions under which they are to be performed. The performance should be described as a visible, overt action to be demonstrated to the assessor. At this point, conditions should identify any special characteristics of the assessment

setting, such as day or night conditions and major types of equipment or equipment settings required for the proficiency assessment. Table 1 lists the nine assessment objectives specified for the *Lookout* assessment. This example illustrates very broad condition descriptions, limited to day, night, and restricted visibility conditions.

Table 1. Summary of *Lookout* assessment objectives.

1. Describe lookout duties and responsibilities.
2. Identify lookout stations and safe routes onboard.
3. Describe and identify the international distress signals.
4. Demonstrate lookout techniques and make lookout reports in clear visibility during daylight.
5. Demonstrate lookout techniques and make lookout reports in clear visibility at night.
6. Demonstrate lookout techniques and make lookout reports in restricted visibility.
7. Demonstrate the use of lookout equipment.
8. Demonstrate man overboard procedures.
9. Demonstrate lookout watch relief procedures.

2.0 Determine Assessment Methods

Goal of this step

Select the most appropriate method for conducting the assessment of each objective.

Sub-steps

- 2.1 Identify alternative assessment methods.
- 2.2 Review the advantages and disadvantages of alternative assessment methods.
- 2.3 Determine assessment methods.

2.1 Identify alternative assessment methods

The intent of the STCW Code is to foster assessment through practical demonstration. Thus, it is important that assessment developers carefully consider the alternative assessment methods available to them and attempt to select the method that best meets the intent of the STCW Code, while also considering the practical limitations of mariner assessment.

This step begins with the consideration of alternative assessment methods for each assessment objective identified in the preceding step. There are three basic methods of assessment:

- Oral or written questions.
- Exercises in a simulated job setting.
- Exercises in an actual shipboard job setting.

Oral or written questions ask the candidate to respond to factual questions regarding general or

job-specific knowledge. Simulated job settings vary widely in the extent to which they accurately present cues from work settings and require candidates to respond in the same manner as onboard ship. Assessments conducted in the actual shipboard job setting provide the candidate with actual work setting cues and require candidate responses that reflect actual job requirements. The reader should note, however, that even in an actual shipboard job setting, the conditions of assessment frequently are modified from non-assessment conditions. For example, at the onset of assessment it might be necessary to provide the candidate with a more explicit explanation of the current situation than is typically provided during normal operations.

At the most general level, there are two basic categories of *what* is being assessed by an assessment procedure, as summarized in Table 2. First, an assessment can evaluate the candidate’s demonstration of *knowledge* of concepts, constructs, rules, and procedures. Often, specific areas of knowledge can be identified as a prerequisite to proficiency in job performance. In such cases, assessment of a candidate’s knowledge is appropriate. Second, an assessment can evaluate the candidate’s demonstration of *skills and ability* to apply knowledge of concepts, constructs, rules, and procedures in an operational setting. Demonstration of ability in operational settings provides a much more *valid* basis of assessment. However, sometimes the assessment of knowledge demonstration in a non-operational setting is more appropriate, due to economic, efficiency, or safety concerns.

The general category of *what* is being assessed corresponds to the general *method* of mariner assessment, as summarized in the right-hand column of Table 2. As presented here, oral or written questions are used to assess the mariner’s demonstration of knowledge. Simulated or actual shipboard job settings are used to assess the mariner’s demonstration of the ability to apply knowledge in an operational setting.

Table 2. General guidance on the selection of assessment methods.

What is Being Assessed	Method of Mariner Assessment
Knowledge of concepts, constructs, rules, and procedures	Oral or written questions
Skills and ability to apply knowledge in an operational setting	Practical demonstration in real or simulated operations

2.2 Review the advantages and disadvantages of alternative assessment methods

Once the alternative assessment methods have been identified, the assessment developer should consider the advantages and disadvantages of each alternative. Factors to consider are assessment validity, assessment reliability, ease of assessment development, and ease of assessment administration. Recall that a *valid* assessment is one that represents critical performance in the operational setting. A *reliable* assessment is one that produces consistent results from assessor to assessor, and candidate to candidate. Table 3 summarizes some of the common advantages and disadvantages associated with each of the three general assessment methods.

Table 3. Assessment method advantages and disadvantages.

Method	Advantage	Disadvantage
Oral or written questions	Easy to standardize Easy to test many at once Good measure of knowledge	May not adequately represent actual requirements of work setting (validity)
Simulator demonstration	High reliability Safe May have high validity	May not adequately represent actual requirements of work setting (validity) May be expensive Usually need one assessor per candidate
Shipboard demonstration	High validity	Usually need one assessor per candidate More difficult to control conditions (reliability) Task may be infrequent or dangerous

2.3 Determine assessment methods

The assessment developer should consider each assessment objective and, in turn, determine an assessment method. The specification of the assessment method should include a general description of the assessment method and assessment conditions for each objective. Assessment methods and conditions must reflect a reasonable level of operational job requirements. The assessment methods must also provide a means of maintaining sufficient control over events so that mariner safety is ensured and operational risks are minimized.

The assessment methods defined for the *Lookout* assessment were initially divided between (1) oral or written questions and (2) practical demonstration in shipboard or simulated operations. This provided flexibility for future applications in the use of either oral or written and either simulator or shipboard assessment. Table 4 summarizes the assessment method selected for each *Lookout* assessment objective from the sample assessment procedure.

Table 4. Lookout assessment objectives and methods.

Assessment Objective	Assessment Method
Describe lookout duties and responsibilities.	Oral or written questions
Identify lookout stations and safe routes onboard.	Oral or written questions
Describe and identify the international distress signals.	Oral or written questions
Demonstrate lookout techniques and make lookout reports in clear visibility during daylight.	Practical demonstration – shipboard and/or ship bridge simulator
Demonstrate lookout techniques and make lookout reports in clear visibility at night.	Practical demonstration – shipboard and/or ship bridge simulator
Demonstrate lookout techniques and make lookout reports in restricted visibility during daylight or at night.	Practical demonstration – shipboard and/or ship bridge simulator
Demonstrate the use of lookout equipment.	Practical demonstration – shipboard and/or ship bridge simulator
Demonstrate man overboard procedures.	Practical demonstration – shipboard and/or ship bridge simulator
Demonstrate lookout watch relief procedures.	Practical demonstration – shipboard and/or ship bridge simulator

3.0 Specify Assessment Conditions

Goal of this step

Outline the assessment conditions for the candidate and assessor.

Sub-steps

- 3.1 Describe assessment settings or scenarios.
- 3.2 Specify oral and written questions.

3.1 Describe assessment settings or scenarios

Following the determination of assessment objectives and methods, the assessment developer should begin preparing a more detailed description of *how* the assessment will be conducted. The assessment conditions provide a good point of departure for this detailed specification. The following should be included in a detailed description of the assessment conditions:

- Assessor instructions.
- Candidate instructions.
- Test conditions, tools, and apparatus used by candidate.

- Specific simulator scenario requirements.
- Requirements for shipboard operations.

The assessor instructions should address special assessment requirements for which the assessor needs to prepare in advance. In the case of simulator-based assessments, this includes familiarization with the controls, displays, and outputs of the simulator. In the case of shipboard assessments, it includes specific operational conditions required to ensure a valid and reliable assessment. For example, if a lookout assessment is being conducted onboard a vessel, it may be necessary to have a specific number of targets visible during the course of the assessment. In addition, issues of personnel safety, operational risk, and operational efficiency should be identified at this point.

Candidate instructions should provide the candidates with the information they will need to prepare for assessment. If simulators are being used, a detailed description of the operational scenario should be provided, specifying both the initial settings of equipment and how the candidate will be allowed to operate controls during the course of the assessment. For shipboard assessment, the operations under which assessment is to be conducted must be described and the initial settings of operational equipment should also be specified. This description requires comparable detail to that for simulator assessment, although the ability of the assessor to control these conditions may be more limited. Table 5 provides a summary of the assessment conditions described for *Lookout* assessment Objective 4.

Table 5. Summary of assessment conditions for *Lookout* assessment Objective 4.

Assessment Objective	Demonstrate lookout techniques and make lookout reports in clear visibility during daylight.
Initial Condition	The assessment should be conducted in clear visibility during daylight. It is essential that reportable objects are in sight.
Required Equipment, Apparatus, and/or Tools	The candidate should be posted at a lookout station equipped with an internal communications system, ship's bell, 7 X 50 individual eye focus binoculars, and bearing repeater fitted with a bearing/azimuth circle, alidade, or peloris. The lookout station should be clear and the assessor must be able to observe activities.

3.2 Specify oral and written questions

Oral and written questions should be adequately specified to ensure comprehensive and consistent assessment. The topics to be addressed by the set of questions should be adequately defined so that this part of the assessment will address a common set of concepts, constructs, rules, or problem-solving skills from one assessment to the next. In the best case, a library of test items will be developed, along with procedures for sampling from this library for any assessment. At a minimum, the number of questions should be identified, as well as the topics and subtopics for subsets of questions. In addition, the type of question (e.g., open-ended, multiple-choice, fill-in, and essay/discussion) should be specified.

Table 6 lists the oral questions included in the *Lookout* assessment, providing more detail on the assessment objectives and methods identified in Table 4. These questions address assessment

Objectives 1, 2, and 3. The question format is open-ended. The candidate is instructed to provide complete answers, and the assessor is asked not to provide coaching or feedback during the questioning period.

Table 6. Oral questions from the *Lookout* assessment.

Assessment Objective	Written or Oral Test Questions
1. Describe lookout duties and responsibilities.	What are the duties and responsibilities of the lookout?
2. Identify lookout stations and safe routes onboard.	Identify three common lookout stations on this vessel. Identify the safe routes to the bow lookout station on this vessel.
3. Describe and identify the international distress signals.	Describe and identify six international distress signals.

4.0 Develop Proficiency Criteria

Goal of this step

Specify the criteria that will be used to assess mariner proficiency, including the measures, standards, proficiency criteria, and scoring procedures.

Sub-steps

- 4.1 Identify components of assessment objectives.
- 4.2 Develop individual measures and standards.
- 4.3 Develop proficiency criteria and scoring procedures.
- 4.4 Validate measures, standards, proficiency criteria, and scoring procedures.

4.1 Identify components of assessment objectives

To this point, assessment objectives have been defined that correspond to basic job requirements identified in the *STCW Code*, other regulatory requirements, or job/task descriptions. The performance required by an objective may need to be broken into smaller, more discrete components for assessment purposes. The separate components of an objective consist of knowledge and/or application of concepts or constructs, rules of operation, or procedures for operation. Separate sources of information are available to assist in the identification of components corresponding to each of these three types.

Three separate information sources are recommended in Table 7 for each type of assessment objective component, with each set including a type of expert. In general, it is recommended that broadly referenced sources (i.e., textbooks, regulations, and technical manuals) be reviewed first. These can be followed by more specific references developed for particular applications (i.e., course materials, company policy, and ship manuals). Finally, experts (i.e., training experts, regulation and policy experts, and operational experts) should be consulted. Information

that is more specific to a course, company, or ship is often required to identify assessment components that realistically reflect job requirements. Frequently, limited reference material will be available, requiring greater reliance on expert input. It is critical to ensure the validity of expert information because there are many perspectives regarding concepts, interpretations of rules, and operationally specific steps and procedures. Assessment should address *commonly accepted* assessment objective components. Therefore, experts should be selected with care, and input from multiple sources should be obtained and compared.

Table 7. Recommended information sources for identifying components of assessment objectives.

Type of Assessment Objective Component	Recommended Information Sources
Concepts or constructs	Textbooks Course materials Training experts
Rules of Operation	Regulations Organizational policies Regulation and policy experts
Steps or Procedures of Operations	Technical manuals Ship manuals Operations experts

For the *Lookout* assessment, three components of assessment Objective 8, *Demonstrate man overboard techniques*, were identified on the basis of a review of ship manuals and interviews with operations experts. These were (1.) throw a ring life buoy, (2.) sound the alarm, and (3.) keep the victim in sight at all times.

4.2 Develop individual measures and standards

Reliable and valid assessment requires the consistent measurement of mariner performance and the consistent application of relevant performance standards. This process can sometimes be relatively complex. As a consequence, the following discussion is the most involved in this manual. The instructions for measure and standard development rely upon the definition of several key terms, as discussed in the following paragraphs.

A **performance measure** is a description of how a candidate’s performance is to be observed and recorded. A useful distinction between types of measures contrasts instrumented and observational measures. An **instrumented measure** relies upon apparatus that can be calibrated to yield highly consistent and accurate measurement results. Examples of apparatus used in instrumented measurement are stopwatches, compasses, rudder angle indicators, oil pressure gauges, and volt meters. An example of an instrumented performance measure is the time required for launching a lifeboat. A complete description of how this measure is to be applied includes (1) how to signal the start of the launching assessment, (2) when to start the timer, (3) the type of timer used and its calibration, and (4) when to stop the timer and record the launch

time. Whenever practical and meaningful, instrumented measures should be used in assessment because they provide the most reliable means of measurement.

An **observational measure** relies primarily on the assessor’s observation and interpretation of mariner performance. In applying an observational measure, an assessor will observe and categorize the mariner’s performance with reference to an established standard. A description of how an observational measure is to be applied is comprised of the recording form and the instructions for how to interpret the observed performance. Although not as inherently consistent and accurate as instrumented measures, the consistency and accuracy of observational measures can be greatly enhanced by carefully describing how the assessor should apply the established standards to the observed performance.

A **performance standard** is the level of a measure that is established as an acceptable or target level. In the case of an instrumented measure, the acceptable level is established independently of the observational procedure. For the example above describing the observation of the time to launch a lifeboat, the standard would be an established time limit that must be met for acceptable performance. In the case of an observational measure, the standard is established as a part of the observational procedure. If the assessor interprets the observed performance as matching the important characteristics of the established standard, the performance is acceptable.

Figure 1 outlines a general procedure for developing measures and standards. The process is linear, with the exception of one branch that must be taken on the basis of the type of performance measure that is used. The following discussion describes each step in this process.

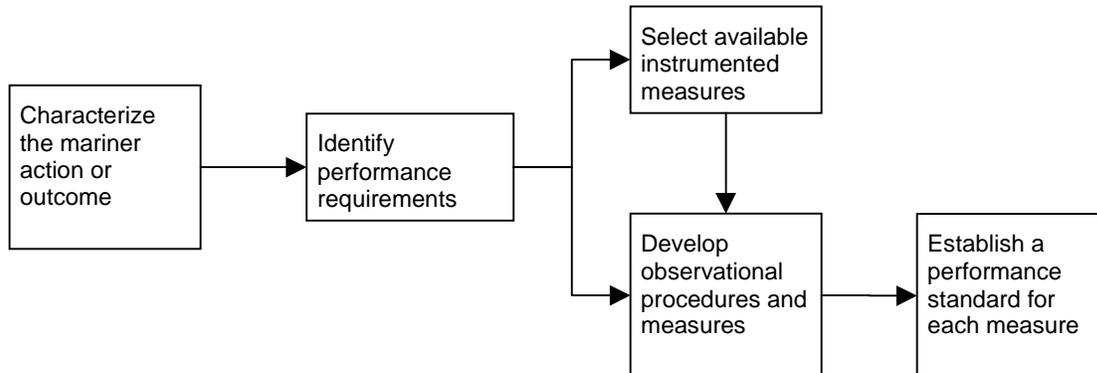


Figure 1. General process for developing measures and standards.

Characterize the mariner action or outcome. The first step in this process is to characterize the actions, or outcomes of actions, performed by the mariner into one of three categories. First, the mariner’s action may be a verbal or written response to a question. Second, the mariner’s action may be the performance of a sequence of operational actions. Third, the mariner’s action may result in an outcome that can be readily observed by the assessor. The characteristics of each mariner action or outcome should be described to allow subsequent consideration of the corresponding performance requirements. Table 8 provides examples of actions and outcomes that are best characterized by one of these three categories.

Table 8. Examples of three categories of mariner actions or outcomes.

Category of Mariner Action or Outcome	Example Descriptions
Verbal or written response to question	Verbal response to question: <i>What are the duties and responsibilities of the lookout?</i>
Sequence of operational actions	Keep the victim (person overboard) in sight at all times
Action resulting in an observable outcome	Report sighted objects verbally using degrees

Identify performance requirements for each action or outcome. The second step is to identify the performance requirements of each mariner action or outcome. Here, a performance requirement is an aspect or characteristic of the action or outcome required for successful job performance. A mariner action or outcome may have a single requirement or it may have several requirements. An example of an action with a single requirement from the *Lookout* assessment is *Call the bridge if you are not properly relieved*. The single requirement is to notify the bridge. An example of an action with multiple requirements for the *Lookout* assessment is *Reports sighted objects verbally using degrees*. This action requires the lookout's reporting (1) a minimum number of objects, (2) what each object is, (3) where each object is, and (4) how far away each object is. During this step, separate performance requirements will be identified. During the next two steps, separate measures corresponding to each requirement will be established.

Select instrumented measures when available. Because instrumented measures are inherently more objective, reliable, and free of observer bias, they should be selected for use in assessments whenever available. In the present *Lookout* example of *Reports sighted objects verbally using degrees*, the use by the assessor of a bearing repeater, or some other device providing accurate bearings, is indicated. In the engine room environment, many actions rely on the reading of calibrated indicators or the operation of calibrated controls. In this case, the instruments are readily available to the assessor to support many of the performance measures.

Develop observational procedures and measures. All performance measures depend upon observation by an assessor. Thus, observational procedures should be specified in all cases where they are not readily understood in advance. If the assessor is using observation based on a reading from an instrument, there will typically be standard procedures for observation. For example, accurate reading of many indicators requires that the observer's eyes be positioned directly above the pointer to avoid the affects of parallax, or apparent displacement. In addition, it will often be necessary to specify the instrumented measure to be used by the assessor. For example, using a bearing repeater to measure performance of *Reports objects verbally using degrees* includes the recording of the deviation, in degrees, between the reported and actual bearing of the target.

Observational measures require special attention in the development of observational procedures. Here, the measure is actually defined on the basis of the explicit observational procedures. In the case of verbal or written questions, this will commonly involve specifying the question to be asked by the assessor. For example, the measure for the *Lookout* action *Identify and describe the international distress signals* consists of the question asked by the assessor *Identify and describe*

six international distress signals. In the case of the observational measure of actions or outcomes, the development of observational procedures requires specifying the individual steps or characteristics of performance. For the *Lookout* example, *Reports sighted objects verbally using degree*, this development involves specifying the assessor's procedure for observing each identified performance requirement. These include (1) counting the number of objects reported, (2) noting the reported and actual type of each object reported, (3) recording the reported and actual bearing of each object, and (4) noting the reported and actual distance of each object.

Establish a performance standard for each measure. Then, it is necessary to establish a performance standard for each measure. The standard is extremely critical to the effectiveness of the assessment in ensuring the proficiency of the mariner. In developing standards for STCW certification, the criterion for proficiency specified in that document is an important first source. Other sources for establishing standards are the same ones used for establishing the assessment objectives and for breaking them into components. These include government regulations, technical manuals, job instructions, textbooks, and task analyses. Finally, appropriate experts can and should be consulted. The standard of performance for each measure should be consistent with important sources and expert opinion. In the case of instrumented measures, this standard should reflect the instrument's unit of measurement. As an example, in the case of the *Lookout* measure of reporting the sighted object's bearing, the standard was established that *Reports must be within 22.5 degrees of the actual bearing of targets.* The experts who established this standard considered the consequences of inaccurate lookout reports. Based on that consideration, it was determined that a report that directed the officer of the watch to a 45 degree area would be adequate to meet the requirement for *a full appraisal of the situation and of the risk of collision* (from Rule 5 of the *Navigation Rules*).

Establishing performance standards for observational measures follows an analogous process to that for instrumented measures. However, in this case, the unit of measure and standard are typically defined together. For oral or written questions, this will require a definition of acceptable answers. For example, for the *Lookout* measure "Answers question *Identify and describe six international distress signals*," the following potential answers are predefined; and the standard of having the candidate correctly recall six distress signals was established on the basis of expert review.

- Red star shells
- Flames on a vessel
- Orange background black ball & square
- "Mayday" by radio
- Dye marker (any color)
- Square flag and ball
- Radio-telephone alarm
- Smoke
- Fog horn continuing sounding
- Gun fired at intervals of 1 minute
- SOS
- Parachute red flare
- Code flags November Charlie
- Wave arms
- Position indicating radio beacon

When the observational measure is based on the assessor's observation of procedural steps or outcomes, it is sometimes much more difficult to predefine all alternative mariner actions. However, when possible, guidance should be provided in the assessment procedure. For example, in establishing the standard for the *Lookout* measure *Report how far off the sighted*

object is, the standard is *correctly reporting using one of four alternative statements (hull-down, on the horizon, hull-up, or close aboard)*. In other cases, the expertise of the assessor must be relied upon to determine if the candidate’s performance meets the standard. For example, the *Lookout* assessment requires correct reporting of *what type of object*. Here, it is left to the assessor’s expert judgment to determine if the lookout’s report provides sufficient information to meet the proficiency criterion as intended in STCW and other sources.

While the ISD literature describes a variety of approaches to developing performance standards, only “pass/fail” standards are described here. During assessment trials, it was determined that the pass/fail approach was the most practical for the assessor doing the over-the-shoulder assessment. In the simplest case, there will be a single pass/fail standard for a single measure and a single action. Table 9 is an excerpt from the *Lookout* assessment that illustrates such a case. Here, if six or more correct international distress signals are recalled, the candidate’s performance meets that standard for Action 3.1, or “passes;” if fewer than six are recalled, the candidate’s performance fails to meet the standard for.

Table 9. Example of a single measure and standard applied to an assessment action.

Action	Performance Measure	Performance Standard
3.1 Identify and describe the international distress signals	Answers question: <i>Identify and describe six international distress signals.</i>	Identifies and describes six of the following: <input type="checkbox"/> Red star shells <input type="checkbox"/> Fog horn continuing sounding <input type="checkbox"/> Flames on a vessel <input type="checkbox"/> Gun fired at intervals of 1 minute <input type="checkbox"/> Orange background black ball & square <input type="checkbox"/> SOS <input type="checkbox"/> “Mayday” by radio <input type="checkbox"/> Parachute red flare <input type="checkbox"/> Dye marker (any color) <input type="checkbox"/> Code flags November Charlie <input type="checkbox"/> Square flag and ball <input type="checkbox"/> Wave arms <input type="checkbox"/> Radio-telephone alarm <input type="checkbox"/> Position indicating radio beacon <input type="checkbox"/> Smoke

When multiple measures and standards are developed for a single action, each of the individual standards must be met to pass the requirements for that action. Table 10 is an excerpt from the *Lookout* assessment that applies this more complex scoring procedure. In this case, the candidate must correctly report (1) a minimum of three objects, (2) what (type of object), (3) where (within 22.5 degrees), and (4) how far off (hull-down, on the horizon, hull-up, or close aboard) to meet

the set of multiple standards.

Table 10. Example of multiple standards applied to a single action.

Action	Performance Measure	Performance Standard
4.3 Report sighted objects verbally using degrees	Reports sighted objects verbally using degrees	<ul style="list-style-type: none"> ❑ Verbally reports at least 3 surface objects. Reports all visible objects. ❑ What (type of objects). ❑ Where (bearings, relative or true). Reports must be within 22.5° of the actual bearing of detected targets. ❑ How far-off (hull-down, on the horizon, hull-up, close aboard). <p>Failure of any of the above four standards constitutes failure.</p>

4.3 Develop proficiency criteria and scoring procedures

Mariner proficiency refers to an individual’s demonstrated ability to meet the job performance requirements. Up to this point in the development of assessment procedures, job requirements have been analyzed to identify individual assessment objectives and these have been further analyzed into component actions. Each action has been reviewed to identify appropriate performance requirements, and individual measures and standards have been developed. For each action, a pass/fail standard has been developed. Now, *proficiency criteria* should be developed for each assessment objective that combine the standards for individual actions on the basis of the action’s *job criticality*. Figure 2 outlines the general process for developing proficiency criteria and scoring procedures. This process involves separate, parallel paths for *critical* and *non-critical* actions, concluding with the development of a combined proficiency criteria and scoring procedure for a single assessment objective. The following discussion describes this process.

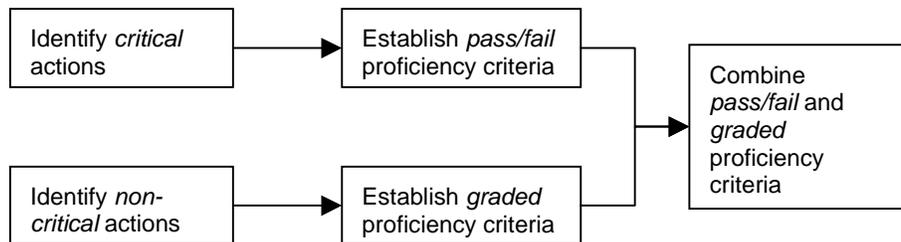


Figure 2. General process for developing proficiency criteria and scoring procedures.

Identify critical actions. The first step in developing proficiency criteria and scoring procedures requires a final consideration of the criticality of the action being assessed. The key consideration is whether the action is essential to safe and/or effective job performance. An example of a critical action is in the *Demonstrate man overboard procedures* objective of the *Lookout* assessment. Successfully meeting the three standards of (1) *establishes and maintains visual contact*, (2) *points in the direction of the person in the water*, and (3) *remains on station until he or she is properly relieved or ordered otherwise* are all considered essential to safety. All actions under a single assessment objective should be considered individually and all critical actions identified.

Establish pass/fail proficiency criteria. Because critical actions are essential for safe and/or effective job performance, a pass/fail proficiency criterion should be established for each of these. That is, failure to meet the standard(s) corresponding to that action will be established as sufficient for failing the proficiency criteria of the corresponding assessment objective. Continuing with the *Lookout* objective *Demonstrate man overboard procedures*, failure to meet the standards for the action *Keep the victim in sight at all times* was established as sufficient for failure of this objective.

Identify non-critical actions. The lower path in Figure 2 is taken for those actions that correspond to non-critical job requirements. Here, meeting the standards for each action is not essential to operational safety and/or effectiveness. For example, everyone is familiar with driver's license tests that have established passing rates, such as 80 percent. In these cases, the assessment typically addresses a procedure that includes safety redundancies or a broad area of knowledge where general, not perfect, understanding and comprehension is established as the target. In the present *Lookout* example, all actions were determined to be critical. However, other, non-critical lookout actions could be identified as part of an assessment of lookout proficiency.

Establish graded proficiency criteria. If a related set of non-critical actions has been identified for an assessment objective, it may be most appropriate to develop one or more sets of graded proficiency criteria. The related set of actions should be reviewed by operational experts, who should assess the risks associated with failure in meeting the standards associated with each action. Then, based on this review, a proficiency criterion should be established for the related set of actions. If more than one set of non-critical, related actions is identified within a single assessment objective, this process should be repeated for each of these sets.

As noted above, the present *Lookout* example does not include non-critical actions. However, subsequent refinement of this assessment for application aboard commercial vessels resulted in the combination of five related *Lookout* assessment objectives. All five of these objectives were judged to involve interrelated knowledge and understanding requirements, and a single proficiency criterion of 80 percent correct was established for the entire set. The reader can refer to these related assessment procedures in McCallum, Barnes, Forsythe, and Smith, 2000.

Combine pass/fail and graded proficiency criteria. The final box in Figure 2 denotes the development of a proficiency criterion across all critical and non-critical actions for the assessment objective under consideration. There are three possible strategies that must be considered at this point, as presented by the hypothetical examples of assessment scores and results in Table 11. If only critical actions have been identified, then a corresponding proficiency criterion can be established that requires the candidate to receive a passing score on

all actions for the corresponding assessment objective. The assessment objective *demonstrate man overboard procedures* is such an example. Here, all three actions were determined to be critical. Thus, the resulting proficiency criterion for this assessment objective requires the candidate to pass the individual standards corresponding to each of the three actions. A hypothetical example of this case is shown in the “All Critical” column of Table 11. For this case, meeting the standards for eight actions corresponds with a score of 8 and a result of not meeting the proficiency criterion for the assessment objective (i.e., failing).

Table 11. Examples of hypothetical assessment scores and results.

Assessment Objective		Criteria, Procedures, & Results	Examples with Different Criticality of Actions			
Action	Score		All Critical	All Non-critical	Critical & Non-critical	
	Pass	Fail				
1.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Proficiency Criteria	Actions 1-10 (critical): 100%	Actions 1-5 (non-critical): 80%	Actions 1-5 (critical) 100%
2.	<input checked="" type="checkbox"/>	<input type="checkbox"/>				
3.	<input checked="" type="checkbox"/>	<input type="checkbox"/>				
4.	<input type="checkbox"/>	<input checked="" type="checkbox"/>				
5.	<input checked="" type="checkbox"/>	<input type="checkbox"/>				
6.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	Scoring Procedures	Actions 1-10 = 80% FAIL	Actions 1-5 = 80% PASS	Actions 1-5 = 80% FAIL
7.	<input type="checkbox"/>	<input checked="" type="checkbox"/>			Actions 6-10 = 80% PASS	Actions 6-10 = 80% PASS
8.	<input checked="" type="checkbox"/>	<input type="checkbox"/>				
9.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Assessment Result	Apply criterion for assessment objective	Apply both criteria for assessment objective	Apply both criteria for assessment objective
10.	<input checked="" type="checkbox"/>	<input type="checkbox"/>		FAIL	PASS	FAIL
	<input checked="" type="checkbox"/>	<input type="checkbox"/>				

When only non-critical actions have been identified, a single graded proficiency criterion or combined set of graded proficiency criteria should be established. If a single set of interrelated actions is identified, a single graded proficiency can be established. However, if analysis of the set of actions reveals two or more independent sets of actions, each with unrelated implications of failure, then two independent sets of graded proficiency criteria should be established. A hypothetical example of the latter case is presented in the “All Non-critical” column of Table 11. For this case, scores of 80 percent for the two sets of Actions 1-5 and Actions 6-10, result in meeting both criteria for the assessment objective.

Finally, when both critical and non-critical actions are identified, then the individual sets of actions must be considered separately in establishing proficiency criteria and scoring procedures. The last column in Table 11 presents this example. For this case, Actions 1-5 are critical and require passing each action; and Actions 6-10 are non-critical, requiring a 60 percent passing rate. Combining these two scoring procedures results in not meeting the proficiency criterion for the assessment objective.

4.4 Validate measures, standards, proficiency criteria, and scoring procedures

The proficiency criteria corresponding to each assessment objective must be validated in a manner that ensures that assessment will be conducted consistently and represent the requirements of the job. Proficiency is, in the end, a judgment that takes into account both the risks associated with sub-optimal performance and the costs of requiring optimal performance of all mariners. A commonly accepted means of establishing the validity of an assessment procedure is to convene a meeting of experts who formally review and reach consensus on the validity of the assessment. Within large organizations, it is often most efficient to establish a standing committee to review and validate new or updated assessment procedures. If this approach is not efficient, it may be best to establish either a standing or ad hoc committee from among several organizations.

Validation of an assessment procedure requires that experts in the relevant proficiency independently review each of the assessment development steps up to this point, discuss their areas of concern, and reach agreement that the final assessment procedure will provide a reliable and valid means of assessing mariner proficiency. The validation process can become time-consuming and inefficient if it is not managed effectively. However, if it is managed effectively and all critical issues are considered and agreed upon, the resulting assessment procedure will have a much greater and more enduring value. Refer to the reference section for texts that address procedures for validation (e.g., Jonassen, Hannum, and Tessmer, 1989; Rothwell and Kazanas, 1998). Basic requirements of an adequate validation process based on committee review are summarized below:

- Assemble a committee of at least three members who are fully qualified and experienced in the area of proficiency under consideration.
- Ask the committee to discuss and/or observe performance of the job and tasks associated with the assessment objectives in both realistic work settings and the proposed assessment conditions.
- Ask each expert to consider and comment independently on the acceptability of the assessment objectives, conditions, measures, proposed standards, and proposed proficiency criteria.
- Ask the committee members to review and discuss issues until agreement among all members of the group has been established.

A final recommended step in validation is trial application of the procedure. Here, draft versions of the assessment are used on a trial basis to identify issues that would not be recognized otherwise. This final step is addressed in the final discussion of implementation issues that follows.

5.0 Prepare the Assessment Materials

Goal of this step

Prepare all materials and tools so that they are ready for use by an assessor and candidate.

Sub-steps

- 5.1 Prepare detailed assessment worksheets.
- 5.2 Prepare assessor instructions.
- 5.3 Prepare candidate instructions.
- 5.4 Finalize written and oral test questions.
- 5.5 Finalize simulator and/or shipboard procedures.
- 5.6 Document references, source materials, and validation process.

5.1 Prepare detailed assessment worksheets

The assessment materials must be prepared as stand-alone procedures that can be administered by assessors who are not familiar with the development process. This will help to yield valid and reliable assessment results. The central assessment materials are the written procedures or worksheets that will be used by the assessor while conducting the assessment. These worksheets should be prepared first, since they will likely be referred to in all of the other assessment materials. The present *Lookout* example has a series of seven worksheets, each of which corresponds to one or more assessment objectives. Each worksheet is formatted into three sections, consisting of (1) identifying information, (2) scoring and recording procedures, and (3) comments that provide succinct instructions for the assessor.

The assessment worksheet should provide sufficient information to guide the assessor through the assessment procedure, as well as a means of recording the outcome corresponding to the individual objectives, objective components, measures, and standards. It should be noted that detailed assessor instructions could take the place of highly elaborate assessment worksheets. That is, much of the information required for conducting an assessment can be included in the assessor instructions, with only the minimum necessary for recording the results of the assessment included in the worksheet. On the other hand, assessor instructions can be limited to the necessary information, with more detailed guidance provided directly on the worksheet. This choice should be based on a consideration of the amount of assessment training available to assessors and the amount of information that can be reasonably included in these two documents, among other considerations.

5.2 Prepare assessor instructions

Every set of assessment materials should include a set of assessor instructions. In many cases, detailed step-by-step instructions help to ensure the reliability and validity of the resulting assessment. Although proficient mariners should be conducting the assessment, there can be many interpretations of how an assessment is to be conducted, especially with regard to candidate instructions, equipment set-up, timing and pace of assessment activities, and interaction between the assessor and candidate.

The assessor instructions could include the following topics:

- Prepare for assessment.
- Conduct pre-briefing with the candidate.

- Observe performance and record results.
- Determine assessment outcome.
- Conduct assessment debriefing.
- Special assessment instructions.

The *Lookout* assessment includes an assessor instructions checklist (see the assessment procedure at the back of this manual) that provides detailed guidance on conducting the *Lookout* assessment, following the topics outlined above.

5.3 Prepare candidate instructions

The set of assessment materials should also include separate, self-contained candidate instructions that can be provided to the candidate well in advance of when the assessment is conducted. At a minimum, the candidate instructions should include:

- Assessment objective source – *STCW Code* chapter, table, function, competence, and proficiency.
- Assessment objective(s).
- Assessment method(s).
- Conditions under which the assessment will be conducted (equipment status, underway versus in port, etc.).
- Assessment procedures, in terms of who will administer the assessment and how long it will take.
- A summary of the measures and standards that will be used to determine the outcome of the assessment.

The *Lookout* assessment includes a set of candidate instructions (see the assessment procedure at the back of this manual) that address what will take place during the pre-assessment briefing, performance observation and recording, and determination of the assessment outcome.

5.4 Finalize oral and written questions

As noted earlier in the discussion of sub-step 3.2 – *Specify oral and written questions*, written items should be documented by either defining detailed question objectives or developing a library of questions. The assessment developer must be aware of the dangers of inadequate specification of oral and written questions. *Without a detailed set of questions, the performance measures are not defined and any performance standards are arbitrary.* At a minimum, the specific knowledge areas to be assessed should be defined, along with the format of the assessment items (i.e., true/false, multiple-choice, short answer, or essay/discussion) and the number of items that should be included for each knowledge area. The developer should consider providing a detailed list of specific oral and/or written questions, along with the correct answers for each question.

5.5 Finalize simulator and/or shipboard procedures

Just as it is necessary to detail written and oral test questions, detailed procedures for simulator and/or shipboard assessment procedures should also be finalized in the assessment materials. When simulators are used, specific assessment scenarios or scripts should be developed. In addition to performance measurement and recording procedures, fully documented simulator scripts or scenarios might include the following specifications:

- Simulated time of day and weather conditions.
- Own vessel status at the start of the scenario.
- Other vessel locations at the start of the scenario.
- Sequence and timing of own vessel malfunctions, if any.
- Course, speed, and maneuvers of other vessels, if any.

In the case of shipboard assessment, a special concern is the need to develop a detailed plan for the assessment. It will be necessary to find a balance between *controlling* and *selecting* the conditions of assessment. The shipboard assessment procedures should detail both what conditions the assessor should establish in advance of the assessment and what conditions must be taken advantage of or avoided when conducting an assessment during normal operations. Documentation of shipboard assessment procedures should include a worksheet that provides space and instructions for the assessor regarding the following items, as applicable:

- Critical safety issues.
- Equipment initialization and set-up.
- Recording of own vessel conditions (date, time, location, course, speed, equipment status).
- Environmental conditions.
- Traffic conditions.
- Cues to be provided to the candidate by the assessor or equipment.
- Performance to be observed and recorded.

5.6 Document references, source materials, and the validation process

The final step in preparing the assessment materials is to document references, source materials, and the validation process. This last step serves two purposes. First, it provides the necessary references and documentation to allow either the assessment developer, or a successor, to maintain and update the assessment procedure. Second, this documentation provides necessary reference material to assist an external reviewer who may evaluate the assessment procedure.

Implementation

When all assessment materials are prepared, the assessment procedure is ready for implementation. Implementation may be as straightforward as handing the document over to the

assessor to begin use in assessing candidates. Or implementation may involve a more complex process that requires several iterations of trial application and refinement before the procedures are ready to serve their function in mariner assessment. More than likely, some intermediate level of trial application and refinement will be required prior to full implementation. Additionally, assessment developers and assessors must remain cognizant of changes in operational equipment, procedures, or regulatory requirements that will necessitate future revisions to the implemented assessment procedures. Issues concerning the implementation of assessment procedures and the conduct of assessments are addressed in a companion manual, *Conducting Mariner Assessments* (McCallum, Barnes, Forsythe, & Smith, 2000).

GLOSSARY

Assessor. Anyone who conducts an *assessment* or evaluation of an individual's *proficiency*. Although the use of the term assessor is retained in some discussions of STCW requirements including NVIC 4-97 on company roles and responsibilities, the term *designated examiner* is used for assessor in the U.S. implementing regulations.

Assessment. The process of evaluating whether an individual's performance meets established *proficiency criteria*. The terminology used for this process in the U.S. implementing regulations includes both an *examination* for knowledge and an *assessment* based on a *practical demonstration* as witnessed by a *designated examiner*.

Assessment conditions. The *assessment conditions* define the setting, tools, references, aids, and safety precautions that are in place at the time that a candidate's proficiency is assessed.

Assessment objective. A critical requirement of job performance that can be measured and assessed. Assessment objectives include skills and knowledge required by the job. A complete *assessment objective* description includes the required mariner performance, the conditions of assessment, and the standards of performance for successful accomplishment of the objective.

Assessment procedures. The activities that are conducted in administering the assessment of a candidate's proficiency. The term *assessment procedure* can be used to describe either the actions taken or the written instructions and activity descriptions that are carried out in conducting an assessment.

Designated examiner. A person who has been trained or instructed in techniques of training or *assessment* and is otherwise qualified to administer performance assessment procedures, in accordance with USCG requirements. Further details on the qualifications of *designated examiner* can be found in NVIC 6-97.

Instrumented measure. A performance measure that relies upon apparatus that can be calibrated to yield highly consistent and accurate measurement results.

Job. A post of employment consisting of a cluster of related work responsibilities and duties (e.g., chief engineer, third mate, able-bodied seaman). In the *STCW Code*, a job is further defined on the basis of licensure level (e.g., officer in charge of a navigational watch on ships of 500 gross tonnage or more).

Knowledge. The learned concepts, cues, facts, rules, and procedures that are necessary for proficient performance of a task (e.g., knowledge of algebra, knowledge of the Navigation Rules, knowledge of procedures for starting the main engine).

Observational measure. A performance measure that relies primarily upon an assessor's observation and interpretation of mariner performance to determine the assessment outcome.

Performance measure. The procedures used by the assessor in observing and recording mariner actions, or the outcome of those actions. The effectiveness of the assessment depends on the care with which these procedures are developed and followed.

Performance standard. The acceptable or target level established for individual *performance*

measures. The assessor must interpret observed performance as passing or failing the established standard.

Proficiency. An individual's demonstrated ability to meet selected job performance requirements, as established on the basis of *performance measures*, *performance standards*, and *proficiency criteria*.

Proficiency criteria. A rule established for combining the measures and standards for individual actions and providing a result for the larger unit of performance encompassed by the assessment objective. Proficiency criteria can be pass/fail to critical actions or graded for non-critical actions.

Qualified instructor. According to the United States implementing regulations, "the person who has been trained or instructed in instructional techniques and is otherwise qualified to provide required training to candidates for licenses, documents, or endorsements." Further details on the qualifications of *qualified instructors* can be found in NVIC 6-97.

Reliability. The *consistency* of a measurement procedure. In the context of assessment, reliability generally can be defined as the consistency of the assessment outcome when applied under comparable conditions. The reliability of an assessment establishes the maximum level of possible assessment *validity*. That is, an assessment can not be any more valid than it is reliable.

Scoring procedure. The defined procedures for combining individual *performance measures* and *performance standards* in the application of *proficiency criteria*.

Skills and abilities. The behaviors that must be applied in successful performance (e.g., typing skills, equipment fault-finding skills, navigation skills, shiphandling skills). Measurable and observable skills are those of interest in proficiency assessment.

Task. A single, observable work assignment that is independent of other actions and results in a valuable outcome. A task must be observable, be a complete work assignment, have a specific beginning and end, and be able to be measured by its intended product or outcome.

Validity. In the context of assessment, validity can be defined as the degree to which an assessment accurately reflects the skill, knowledge, and performance requirements of the job. The maximum validity of an assessment is established on the basis of its *reliability*. That is, an assessment can not be any more valid than it is reliable.

REFERENCES

- 46 CFR §10, 12, and 15. (1997). Implementation of the 1995 Amendments to the International Convention on Standards of Training, Certification and Watchkeeping for Seafarers, 1978 (STCW). U.S. Code of Federal Regulations. Coast Guard, U.S. Department of Transportation. Washington, DC: Office of the Federal Register.
- Gagné, R., Briggs, L., & Wagner, W. (1992). Principles of instructional design (4th ed.). Fort Worth, Texas: Harcourt Brace College Publishers.
- International Maritime Organization. (1996). International Convention on Standards of Training, Certification and Watchkeeping for Seafarers, 1978, as amended in 1995 (STCW Convention). London: Author.
- International Maritime Organization. (1996). Seafarers' training, certification, and watchkeeping (STCW) code. London: Author.
- International Maritime Organization. (1998). Guidance on shipboard assessments of proficiency. Marine Safety Council Circular 853, 22 May, 1998. London: Author.
- Jonassen, D., Hannum, W., & Tessmer, M. (1989). Handbook of task analysis procedures. New York: Praeger.
- Krieger, G. (1994). Constructing and validating tests. In W. Tracey (ed.), Human Resources Management and Development Handbook (3rd ed.). New York: AMACOM.
- McCallum, M.C., Barnes, A.E., Forsythe, A.M., & Smith, M.W. (2000). Conducting Mariner Assessments: A Trial Application of Procedures, Training and Manuals. (Report No. R&DC 204). Groton, CT: U.S. Coast Guard Research and Development Center In development.
- McCallum, M. C., Forsythe, A. M., Smith, M. W., Nunnenkamp, J. M., & Sandberg, G. R. (1999). Developing performance-based assessments of mariner proficiency (Report No. R&DC 303). Groton, CT: U.S. Coast Guard Research and Development Center. In revision.
- Mager, R. (1997). Preparing instructional objectives (3rd ed.). Atlanta, GA: Center for Effective Performance.
- Murphy, J. (1997). Deck officer study guide, volume 1, deck general. Plympton, MA: Academy Publishing Company.
- Rothwell, W. & Kazanas, H. (1998). Mastering the instructional design process (2nd ed.). San Francisco: Jossey-Bass.
- Smith, M. W., Sandberg, G. R., McCallum, M. C., Stewart, R. D., Hard, D. A., Meurn, R. J., & Baker, D.R. (1996). An interactive test of mariner competence. (Report No. CG-D-04-97) Groton, CT: U.S. Coast Guard Research and Development Center. (NTIS AD-A 322240).
- Sullivan, R. & Elenburg, M. (1988). Performance testing for technical trainers. Training and Development Journal. 42 (11), 38-40.

U.S. Coast Guard (1997). Navigation and Vessel Inspection Circular No. 4-97: Guidance on company roles and responsibilities under the 1995 Amendments of the International Convention on Standards of Training, Certification and Watchkeeping (COMDTPUB P16700.4 NVIC 4-97). Washington, DC: Author.

U.S. Coast Guard (1997). Navigation and Vessel Inspection Circular No. 6-97: Policy on qualified instructors and designated examiners who train or assess the competence of merchant mariners (COMDTPUB P16700.4 NVIC 6-97). Washington, DC: Author.

U.S. Coast Guard (1997). Navigation and Vessel Inspection Circular No. 7-97: Guidance on STCW quality standards systems (QSS) for merchant mariner courses or training programs (COMDTPUB P16700.4 NVIC 7-97). Washington, DC: Author.

U.S. Coast Guard, Department of Transportation (1990). Navigation Rules, International-Inland. (COMDTINST M16672.2B). Washington, DC: U.S. Government Printing Office.

[This page left blank intentionally.]