

STANDARD OPERATING PROCEDURES (SOP)
FOR
THE COAST GUARD'S TRAINING SYSTEM

Volume 10

Testing



**Coast Guard Force Readiness Command
April 2015**

THIS PAGE LEFT INTENTIONALLY BLANK

TABLE OF CONTENTS

TITLE	PAGE
Chapter 1 – Testing Process	
<u>Overview</u>	1-1
<u>Participants in the Testing Process</u>	1-3
Chapter 2 – Test Development Procedures	
<u>Testing Philosophy</u>	2-1
<u>Types of Tests</u>	2-3
<u>Criticality of Tests</u>	2-6
<u>Test Construction</u>	2-9
<u>Defensibility of Tests</u>	2-18
Chapter 3 – Determining Test Cut/passing Scores	
<u>Overview of the Angoff Method</u>	3-1
<u>Applying the Modified Angoff Method</u>	3-2
Chapter 4 – Security and Compromise	
<u>Background</u>	4-1
<u>Security of Test Materials</u>	4-4
<u>Compromise</u>	4-15
Chapter 5 – Test Item Writing	
<u>General Principles</u>	5-1
<u>Types of Test Items</u>	5-4
<u>Item Writing Standards</u>	5-6
<u>Figures, Photos, and Graphics</u>	5-40
Chapter 6 – Enlisted Advancement	
<u>Background</u>	6-1
<u>Rating Advancement Test (RAT)</u>	6-4
<u>Servicewide Examination (SWE)</u>	6-22
Chapter 7 – Performance Tests	
<u>Performance Test Development</u>	7-1
<u>Introduction to Performance Test Development</u>	7-2
<u>Collection of Documentation</u>	7-3
<u>Designing Performance Test Items</u>	7-5
<u>Preparing Checklist for Process and Product Evaluation</u>	7-9
<u>Scoring Procedures for Performance Evaluations/Tests</u>	7-12
<u>Writing Performance Test Instructions</u>	7-17

TABLE OF CONTENTS

TITLE	PAGE
Chapter 8 – Analytics	
<u>Overview</u>	8-1
<u>Frequency Distribution</u>	8-2
<u>Commonly Reported Statistics</u>	8-5
 APPENDIX A – Job Aid for Test Design of the Paper Versions of the RAT Using QMP	
<u>Job Aid</u>	A-1
 APPENDIX B – Job Aid for Test Design of the Online Version of the RAT Using QMP	
<u>Job Aid</u>	B-1
 APPENDIX C – Job Aid for the Modified Angoff Method	
<u>Applying the Modified Angoff Method</u>	C-1
<u>Notes Concerning the Modified Angoff Method</u>	C-8
 APPENDIX D – Test Plan	
<u>Sample Test Plan</u>	D-1
 APPENDIX E – Nondisclosure Statement	
<u>Sample Nondisclosure Statement</u>	E-1
 APPENDIX F – Cognitive Levels of Testing	
<u>Cognitive Levels of Testing</u>	F-1
 APPENDIX G – Use of Correct Grammar and Effective English	
<u>Correct Use of Grammar</u>	G-1
<u>Use of Effective English</u>	G-9
 APPENDIX H – Developing Pre- and Posttests	
<u>Pre-Posttest Explained</u>	H-1
<u>Pre-Posttest Cautions</u>	H-2
<u>Use of Pre-Posttest</u>	H-3
<u>Types of Tests</u>	H-4
 APPENDIX I – Sample RAT Reference Web Page	
<u>Web Page Format</u>	I-1

[REFERENCES](#)

Chapter 1

TESTING PROCESS

Overview

Introduction

“Effective test development requires a systematic, well-organized approach to ensure sufficient validity evidence to support the proposed inferences from test scores.”¹

Testing is the collection of quantitative (numerical) information about the degree to which a competence or ability is present in the test taker. There are right and wrong answers to the items on a test, whether it be a test comprised of written questions or a performance test requiring the demonstration of a skill. (Shrock, S. & Coscarelli, W., 2007)

The United States Coast Guard uses testing in a variety of situations and settings to verify and document that military members, civilian employees, merchant mariners and, occasionally, contract personnel meet specific standards. These standards vary and can range from simple verification of knowledge retention to actual performance of specific tasks.

Source Citations

Much of the information in this SOP has been derived from collaboration with the directors of the U.S. Air Force Airman Education and Advancement Center, U.S. Navy Advancement Center, information from Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 2014), and International Standard ISO 10667; Assessment service delivery – Procedures and methods to assess people in work and organizational settings part 1 (Requirements for the client) and part 2 (Requirements for service providers), as well as professional test development standards and processes.

¹ Steven M. Dowling, University of Illinois at Chicago

Overview (continued)

Purpose

The purpose of this volume of the US Coast Guard Training System Standard Operating Procedures (TRASYS SOP) is to provide:

- procedures for the development of test items and tests
- procedures for the evaluation of tests and test items
- guidelines to maintain defensibility of the testing system

By developing test items and tests using standard, industry accepted methods and procedures, the USCG will be able to meet the requirements of ISO 10667-1 (page V), which include:

- defining good practice for assessment procedures and methods;
- ensuring equity in the application of assessment procedures;
- enabling appropriate evaluation of the quality of assessment service provision.

Action

Design, development, administration and analysis of all tests used to measure the competence or ability of the personnel completing FORCECOM managed training as well as advancement of enlisted ratings shall follow the procedures promulgated in this volume of the TRASYS SOP. Requests for deviation from the standards provided must be requested, in writing, with justification, from the U.S. Coast Guard Force Readiness Command (FC-T).

Participants in the Testing Process

Roles of Participants

There are many participants in the Coast Guard testing process, such as those who:

- prepare and develop the test items and tests
- review, edit, and approve the test items and tests
- schedule and administer the tests
- monitor the tests
- score the tests
- make use of test results to make decisions
- interpret test and test item performance statistics
- take the test by choice or by direction

The roles identified above are sometimes combined and may be co-located or assigned to diverse locations. In order to maintain an effective, defensible testing system, it is imperative that personnel filling any role in the testing process communicate effectively with other participants in the process. Effective testing requires that all participants in the testing process possess the knowledge, skills, and abilities relevant to their assigned roles in the testing process. All participants should be aware of personnel and contextual factors that may affect the testing process. As an example, test developers and those selecting and interpreting test results must have adequate knowledge of psychometric principles, especially the concepts of validity and reliability.

Participants in the Testing Process (continued)

Major Roles

In order to ensure smooth operation of the testing process, the US Coast Guard has identified several major roles. These roles should be assigned at each training center as primary duties to be most effective, but may be assigned as collateral functions of existing billets. Multiple roles may be fulfilled by a single person due to staffing and budgetary constraints. The major roles in the USCG testing process are as follows:

Test Development Manager

The Test Development Manager (TDM) monitors Rating Performance Qualifications (RPQs) for assigned U.S. Coast Guard enlisted specialties, analyzes any changes in Servicewide Examinations (SWEs) and Rating Advancement Tests (RATs) development, and incorporates required adjustments into the test development process. In addition, the manager supports resident course knowledge and performance test development as needed for each training center. The TDM should be familiar with test item and test construction as well as psychometric evaluation of tests and individual test items.

Duties of the test development manager are as follows:

- Provides overall test quality control as a reviewer of work submitted by test authors and test item contributors
 - Serves as final authority for the release of enlisted advancement tests
 - Provides training concerning test item development and test item/test analysis as required
 - Acts as TRACEN liaison with Coast Guard Institute (CGI) and Pay & Personnel Center (PPC-adv) on all matters concerning enlisted advancement testing
-

Enterprise Management of Testing Software / Systems

As the designated manager and program sponsor for testing software and systems for the Training Division, FC-Tadl works with stakeholders to collect and validate requirements, engages CG-6 to identify a C4&IT software or system solution that meets program needs, and oversees systems lifecycle continuity and capability on behalf of the program.

Roles and responsibilities of FC-Tadl include:

- Collect and validate software and systems requirements.
- Define, maintain, and articulate program software/system requirements using the Coast Guard's Systems Development Lifecycle (SDLC) process.
- Ensure software / systems have Authority to Operate (ATO) and are integrated within the Coast Guard's enterprise architecture (EA) prior to launch.
- Identify alternative solutions to meet program requirements.

Participants in the Testing Process (continued)

- Identify, communicate, and work to resolve existing systems and software degradation.
-

Rating Knowledge Manager (RKM)

Master Chief Petty Officers have superior leadership, military characteristics, technical knowledge, and duty performance. Rating Knowledge Managers (RKMs) are assigned to each enlisted rating as the primary resident and nonresident training advocate for their rating specialty as prescribed by the Rating Force Master Chief (RFMC) and responsible CGHQ program manager.

Duties of the RKM, relating to test development, are as follows:

- Subject matter expert (SME) for their rating
 - Oversees the Assistant Rating Knowledge Manager (ARKM) in the development of all Rating Advancement Test (RAT) test items, testing instruments, as well as Servicewide (SWE) test items and testing instruments as directed by local command
 - Ensures material presented in the rating specialty “A” school supports each RPQ
 - Maintains the Servicewide Exam (SWE) test database, including reference revisions; serves in the role of author to develop SWE test items and testing instruments
 - Develops the SWE for the May and November active duty and the October reserve forces test administrations
 - Reviews and responds to all challenges to the validity of SWE test items from test takers
 - Reviews the statistical analysis of each administration of the SWE and adjusts the SWE content as required to achieve a wide distribution of scores, as desired by the test plan, if used. Adjusts difficulty and content areas as necessary
 - Reviews the statistical analysis of the RAT with the ARKM semiannually to ensure items are performing, as desired in the test plan, if used. Adjusts as necessary
 - Ensures all rating advancement test items are technically accurate and validated to match the current RPQs and references
 - Serves as chairman/lead facilitator of the Angoff rating group to establish valid cut (pass) scores for the RAT for each paygrade within his/her specialty
-

Participants in the Testing Process (continued)

Assistant Rating Knowledge Manager (ARKM)

The Assistant Rating Knowledge Manager (ARKM) replaces the legacy Subject Matter Specialist (SMS) role for each enlisted rating specialty. The ARKM collects, organizes, and disseminates learning resources in support of the RPQs, to maintain RPQs and associated test-items, and to capture best-practices to improve training and performance support. The ARKM works under the direction of the RKM.

Duties of the ARKM, relating to test development, are as follows:

- Subject matter expert for their rating
- Maintains an online RAT reference library in an up-to-date condition
- Maintains the RAT test item database, including reference revisions
- Serves in the role of author to develop RAT items and testing instruments
- Reviews and responds to all challenges to the validity of RAT items from test takers
- Reviews the statistical analysis of the RAT with the RKM semiannually to ensure items are performing as forecast during the Angoff review and test plan, if used, and adjusts as necessary
- Ensures all rating advancement test items are technically accurate and validated to match the current RPQs and references
- Assists RKM and coordinates the Angoff rating group to establish valid cut (pass) scores for the RAT for each paygrade within his/her specialty
- Assists in development of Servicewide (SWE) test items and testing instruments as directed by local command

Pay & Personnel Center (PPC) Advancements Branch (ADV)

Commanding Officer, (CG PPC) is the single point of contact for all SWE inquiries, corrections, and waivers; and is responsible for the preparation, printing, distribution, accountability and scoring of the ServiceWide examinations. Following the scoring process, the commanding officer is responsible for preparation and distribution of the advancement eligibility lists approved by Commander (CG PSC).

Duties of PPC Advancements Branch (ADV) for Servicewide exam administration is as follows:

- Receive master copies of SWE booklets and answer keys from RKM's for each rating by the required deadline for the three annual SWE cycles
- Page check booklets for correct format/layout/print quality, conferring with RKM's on changes

Participants in the Testing Process (continued)

Pay & Personnel Center (PPC) Advancements Branch (ADV) (continued)

- Transfer answer key data from paper answer keys to the SWE database for future SWE scanning and scoring
- Program the SWE Eligibility Criteria for each rating and each cycle into the SWE database to validate each members advancement eligibility
- Determine the number of eligible candidates for each SWE cycle
- Produce and publish Personal Data Extracts for all enlisted members currently in grades E-4 through E-8 for verification/correction for each SWE cycle
- Receive and act on requests from members and units to update/correct PDE's during published correction periods
- Receive and make determination on all SWE Waiver requests
- Have exam booklets commercially printed based on numbers of eligible candidates
- Package and ship exam booklets, answer sheets, and SWE Instructions to Exam Board units for SWE administration
- Receive SWE Answer Sheets and challenges from Exam Boards
- Scan/score exams to produce raw scores in members Test Results page in DA
- Record, sort and forward challenged questions forms to RKM's for review and respond to PPC(ADV)
- Modify SWE Answer Keys based on RKM response, then rescore all exams
- Compute SWE Standard Score based on population statistics (Convert raw score to standard score)
- Compute member final multiple score based on all points (Standard Score, TIS, TIR, Awards, Sea Points, Surf Points, EERs)
- Rank order members on SWE Boards based on final multiple score then produce and release rank ordered advancement eligibility lists for each rating
- Create and release members personal Profile Letters showing final multiple score breakdown
- Create and provide SWE statistical reports to RKM's as an aid in validating SWE question accuracy and strength
- Maintain eligibility lists and complete monthly advancements based on EPAA/ERAA messages.

Participants in the Testing Process (continued)

Coast Guard Institute (CG Institute)

CG Institute verifies and implements policies, requirements and procedures for all Rating Advancement Test (RAT) and End of Course Test (EOCT), in addition is responsible for publishing and distributing test and testing resources to Education Service Officers throughout the Coast Guard.

- Manage the review, printing and distribution of all RAT and EOCT.
- Maintain an accurate repository of all RAT, EOCT and RAT Reference CD's for the Coast Guard.
- Process the scoring of all RAT and EOCT and upload to Direct Access as official scores.
- Process and archive all RAT and EOCT question challenges.
- Provide guidance for all potential test compromise situations.
- Provide waiver policy guidance and authority for all RAT and EOCT.
- Oversee Defense Language Proficiency (DLPT) & Armed Forces Classification Test (AFCT) administrator access.

Testing System Administrator and Alternate

Each TRACEN shall establish a testing system administrator and alternate testing system administrator. This is a highly sensitive position and should only be filled by the most trusted individuals. The persons filling these roles may have access to all topics, test items, assessments, internal and shared content and resources, schedules, participants, groups, and reports assigned to his or her TRACEN, as well as other TRACENs that are using the system. Ideally, the testing system administrator and alternate should be a civilian GS employee at a paygrade of GS-9 or greater and in no case shall ever be at a paygrade or position below that of the highest enlisted paygrade (E-9) that maintains test items in the database. The testing system administrator shall be well versed in test item construction, the functionality of the current USCG enterprise testing repository. The testing system administrator and alternate shall be guided by the e-testing instructions published and maintained by FC-Tadl.

Primary duties of the testing system administrator and alternate include:

- Serves as the point(s) of contact between the TRACEN and the enterprise testing repository program office (FC-Tadl)
- Coordinates the use of the enterprise testing repository locally and is the approving and relinquishing authority for local installations and addition/deletion of administrators (users) at the assigned TRACEN
- Conducts local enterprise testing repository training for users
- The testing system administrator will normally fill a dual role as the Test Development Manager and serve as the ISS approver in the

Participants in the Testing Process (continued)

**Testing
System
Administrator
and Alternate
(continued)**

enterprise testing repository workflow and may also serve as writer/editor if required

- Develops and maintains a policy for local implementation of the enterprise testing repository and submits to FC-Tadl, via the chain of command, for review
-

Participants in the Testing Process (continued)

Test Item/Test Instrument Authors

Authors are responsible for the development of test items within designated question pools following accepted industry standards and Coast Guard SOPs. When utilizing the CG enterprise testing repository, all test item topics shall be structured as shown by Figure 6-1 in Chapter 6 of this SOP. The test developer will provide 3 to 5 questions per enabling/lesson objective or EPQ/RPQ task when possible. An author shall have access only to those test items topics he or she is responsible for and will follow a locally approved workflow process. In the case of Servicewide Examination (SWE) test items, the author shall be at a paygrade/rank as directed by local command, a civilian GS-9 or above, or a contractor holding an equivalent position.

Primary duties of authors include:

- Developing valid and reliable test items based upon appropriate references and course objectives
- Using appropriate types of test items to achieve desired test of performance, application of knowledge or retention of knowledge
- Developing test instruments, with the assistance of the Test Development Manager, to ensure the proper distribution of test items is maintained as outlined in the test plan (if used)

Education Services Officer (ESO)

In the case of the administration of high stakes examinations, the Educational Service Officer, ESO, is an integral member of the testing process. The ESO shall be guided in his or her duties by current policy set forth by the Coast Guard Institute (CGI) for the administration of RAT and Coast Guard Personnel Services Center Advancement Division (PSC-ADV) for administration of the SWE.

This page intentionally left blank

Chapter 2

TEST DEVELOPMENT PROCEDURES

Testing Philosophy

Purpose of Testing

All tests should be well developed and testing practices, beneficial. There is extensive evidence documenting the effectiveness of well constructed tests in relation to supporting the validity of the test. The proper use of tests can result in making wiser decisions about individuals and programs than those made without using tests. The improper use of tests, however, can cause considerable harm to test takers and others affected by test-based decisions (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1999). Tests can also enable a path to provide a more equitable method to advancement and retention decisions.

Intent

The intent of this section is to provide standards that will promote sound, fair, ethical, and unbiased use of tests and construction of test items. Additionally, adherence to the standards will provide a defensible U.S. Coast Guard testing system.

Cognitive levels of testing are integral to test development and are included in this SOP in Appendix F.

Derivation of Content

Content in this section has been compiled from best practices of U.S. Coast Guard Forces Command, U.S. Air Force Airman Education and Advancement Center, U.S. Navy Advancement Center, U.S. Army Training and Doctrine Command, the Standards for Educational and Psychological Testing, ISO 10667 (Assessment service delivery — Procedures and methods to assess people in work and organizational settings), as well as other industry accepted practices.

What is Testing?

Shrock & Coscarelli (2007) state the following concerning testing:

Testing is the collection of quantitative (numerical) information about the degree to which a competence or ability is present in the test taker. There are right and wrong answers to items on a test, whether it be a test comprised of written questions or a performance test requiring the demonstration of a skill. (p. 15)

The Coast Guard uses testing in several forms, written knowledge, performance-based and actual performance tests, to ensure our members are ready to perform their assigned duties. Testing does not stop at the school – our members are tested every day by the eyes of the public. We must be sure we are mission ready.

Testing Philosophy (continued)

Qualities of a Successful Test

For a test to be successful and defensible, it must meet the following criteria:

- Be based on knowledge, skills, abilities, and performance expected on the job
 - Contain items that require the candidate to apply job knowledge and experience to select the correct alternative
 - Contain items that will demonstrate satisfactory performance regarding construct, content, rational/logical, and face validity. (Unclear, trivial, basic job qualification and rote memory type items should not be used)
 - Meet professionally recognized statistical performance expectations
 - Contain items that are suitable for the intended purpose of the test
 - A test item in a norm-referenced test (NRT), such as the SWE, that is answered correctly by about 62 percent of the population is a widely accepted standard, but there must be a mix of all difficulties to ensure a widely dispersed range of scores. Besides difficulty, there are several other statistical values to consider. These are addressed in Chapter 8 (Analytics) of this SOP.
 - Test items in a criterion-referenced test (CRT), such as the RAT and “A” and “C” school tests/quizzes, should match the requirements of the occupational or job task analysis (OA/JTA) as well as the related course objectives. The degree of success should match the judgment of the Angoff scoring panel (see Chapter 3 of this SOP (Determining Test Cut/Passing Scores). In a CRT, the issue of substitutability must be considered also. This topic is covered in more detail in Chapter 5 (Test Item Writing) of this SOP.
 - Contain items that demonstrate satisfactory content and performance regarding predictive, concurrent, and congruent validity
 - Contain items that are not biased or based upon demographic or geographic specific criteria
-

Types of Tests

Types of Test Forms

Generally, there are two specific types of test forms: norm-referenced and criterion-referenced. Each type has a specific purpose with major philosophical differences in the interpretation of the test scores. Tests must be constructed in order to facilitate their intended purpose of either their norm-referenced or their criterion-referenced interpretation.

Basically, norm-referenced tests are designed to separate test takers from one another, while criterion-referenced tests must be composed of items based upon specific objectives or competency statements.

Within each of these general categories the tests are further divided by criticality, that is, high-stakes, medium stakes, or low stakes. This chapter will provide definitions and examples of each and describe appropriate usage guidelines.

For a discussion of Pre and Post-tests, refer to Appendix H of this SOP.

Norm-Referenced Test

Specific qualities of a norm-referenced test (NRT) include:

- Does not normally have a “cut” or passing score
 - Compares people in relation to the test performance of one another
 - Composed of items that will separate the scores of test-takers from one another
 - Used to rank-order and separate top performers from a pool of those who are already considered at least “minimally qualified”
 - Does not verify what a test taker can actually ‘do’ on the job
 - Test scores should be widespread as this increases confidence for comparison
 - Usually have to be completed within a time limit
 - One more test item, right or wrong, can cause a large change in a student’s final rank or standing
-

Types of NRTs

The Coast Guard utilizes a limited assortment of NRTs to ensure the most qualified personnel are advanced or assigned to specific rating specialties. Examples of Coast Guard NRTs are:

- Armed Services Vocational Aptitude Battery (ASVAB)
 - Measures general knowledge of several vocational areas
 - Has an established ‘cut’ score based upon psychometric studies of desired levels of knowledge of specific areas
 - Used to determine what occupations or specialties candidates are best suited for
-

Types of Tests (continued)

Types of NRTs (continued)

- Servicewide Examination (SWE)
 - Used to rank-order enlisted personnel for advancement to next paygrade
 - Identifies the most vs. least knowledgeable of the entire rating of those who are already minimally qualified for advancement to the next paygrade
 - As demonstrated by previous criterion-referenced testing, performance, time in service, time in grade and other established advancement prerequisites

Some examples of commercial, national, norm-referenced tests include the California Achievement Test (CAT); Comprehensive Test of Basic Skills (CTBS); Iowa Test of Basic Skills (ITBS) and Tests of Academic Proficiency (TAP); Metropolitan Achievement Test (MAT); and Stanford Achievement Test (SAT).

Criterion-Referenced Test

Specific qualities of a criterion-referenced test (CRT) include:

- Compares people against a standard
 - Is composed of items based upon specific objectives or competency statements
 - Defines the performance of each test-taker without regard to the performance of others
 - Defines success as being able to perform specific tasks or competencies
 - Tests the mastery of a concept
 - May be assigned a “cut-score” or passing score by a panel of expert judges (medium to high stakes)
-

Types of Tests (continued)

Types of CRTs

The Coast Guard utilizes an assortment of CRTs to ensure the transference of knowledge to application of actual performance has met minimum acceptable criteria. Some examples of Coast Guard CRTs are:

- Resident course
 - These assessments may be either computer-based or presented on paper
 - Progress quizzes used primarily as student self-checks of comprehension of classroom presentations
 - Section review tests to check student retention of knowledge or application of knowledge
 - Final examination to assess comprehension and application of all skills/knowledge gained by course attendance
 - Role play
 - Tests the application of knowledge in a simulation of actual required on-the-job performance
 - Performance Test Checklist (PTC)
 - Typically used in resident training environments and known as “Go-No-Go” tests. An instructor observes the performance of a student or group of students and compares the actual performance to the required performance shown on a criterion-referenced checklist
 - Rating Performance Qualification Standards (RPQS) completion sheets
 - Rating Advancement Test (RAT)
 - Used to test the mastery of the application of knowledge upon completion of RPQs. The test is written as closely to the job performance as possible and selected references may be provided to the test taker
 - The RAT may be delivered electronically or by paper. A paper-based test with three versions must be created. See Chapter 6 for more details.

CRT vs. NRT

The key differences between the two general test types are:

- Criterion-referenced tests are used to test the mastery of a concept
 - Test scores should be high if the training material and tests are performing their intended purpose
 - Norm-referenced tests are used to rank-order (rack-and-stack) personnel who are already minimally qualified
 - Test scores should be spread evenly along the spectrum
-

Criticality of Tests

Categories of Tests

Tests are further divided into categories based upon the criticality of the results and purpose of the test. Different security procedures are required to prevent compromise and cheating. (Shepherd, E., 1995)The ratings are:

- Low stakes
 - Medium stakes
 - High stakes
-

Low Stakes

A low stakes test is distinguished by the following characteristics:

- Few consequences to test taker for failure
- Low motivation to cheat
- Proctoring not normally required
- Minor amount of development effort required
- Grades may or may not be assigned

Examples of low stakes tests include:

- Instruments to motivate learning
 - Tests that guide students to knowledge and learning resources
 - Self-assessments
 - Tests that promote thinking
 - Instruments that allow for skills gap analysis
-

Medium Stakes

A medium stakes test is distinguished by the following characteristics:

- Some consequences to test taker for failure
- Medium motivation to cheat
- Proctoring sometimes required
- Medium amount of development effort required
- Grades are assigned

Examples of medium stakes tests include:

- Instruments to determine levels of knowledge and skills
 - Practice tests
 - Tests used to determine skill pools and knowledge inventories
 - Instruments that allow for skills gap analysis
-

Criticality of Tests (cont.)

High Stakes

A high stakes test is distinguished by the following characteristics:

- Major consequences to test taker for failure
- High motivation to cheat
- Requires constant proctoring
- Major amount of development effort required
- Grades are assigned

Examples of high stakes test include:

- Tests used for regulatory certifications and licenses
 - Tests used to qualify for regulatory certifications
 - Entrance exams
 - Pre-employment tests
 - Results critical to granting permissions
 - Instruments allowing for skills gap analysis
-

Criticality of Tests (cont.)

Criticality Within the USCG

Within the Coast Guard testing system most tests fall within the low to medium stakes range of criticality. We do have some very important high stakes tests which must be developed, handled, administered, and reviewed carefully. Typically, the rule of thumb is that if the results of the test bear directly upon advancement, accreditation, or certification, the test is considered to be high stakes.

Although this list is not all inclusive, some examples of the various criticalities of tests used in the Coast Guard are as follow:

- Low Stakes:
 - Course pre-test/post-test
 - “A” school progress quiz
- Medium Stakes:
 - Performance Test Checklist
 - School/Course final exam
 - STAN 2.0 written assessments
- High Stakes:
 - Rating Advancement Test (RAT)
 - Servicewide Examination (SWE)
 - Law Enforcement Certification Exams
 - Merchant Mariner Licensing Exams
 - FAA/Aviation Certification/Licensing Exams
 - Enlisted Professional Military Education (EMPE)/Advanced Qualification Exam (AQE)

Test Construction

Test Design

Test design is not a simple task and should not be taken lightly. Shrock and Coscarelli (2007) suggest a process that should be followed when designing criterion-referenced tests. This process can be used effectively for norm-referenced test design, as well, simply by skipping the step to set a cut score. A diagram adapted from their method is presented below as figure 2-1. The most important point is to document the entire process as you proceed to maintain defensibility. This may be done with a test plan. The contents of a test plan is presented later in this chapter and a sample is presented in Appendix D.

Designing Criterion-Referenced Tests

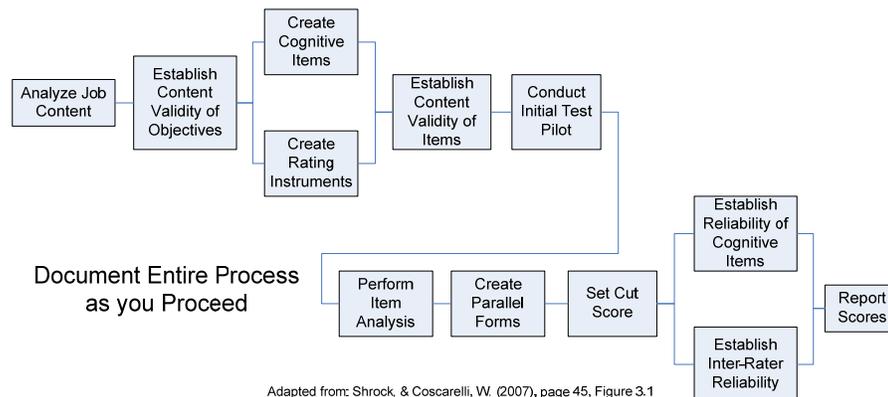


Figure 2-1

Test Length

Deciding how many items should be on a test is not a simple task with a fixed numerical answer. Studies have indicated that the length of a test has a direct relationship with the test's reliability and, therefore, its validity. (Shrock, S. & Coscarelli, W. 2007)

Four factors affect the length of the test:

- How critical are decisions based upon the results of the test?
- What resources (time and money) are available for testing?
- How big is/are the overall objective(s) that is/are being assessed?
- How closely related are the objectives that are being tested?

The simple answer is obvious, the more test items on the test, the greater the reliability. Creating a test with hundreds or thousands of test items is not practical so we must use a general rule of thumb. Each of the four factors above will be addressed briefly.

Test Construction (continued)

Criticality of Test Results

Deciding the criticality of flawed results to both the individual and the Coast Guard is the first step in deciding test length. Questions to ask include:

- What are the consequences to the Coast Guard of erroneously allowing a non-master (minimal performer) to pass a test?
 - Undeserved advancement?
 - Poor work performance?
 - Damage to equipment?
 - Lawsuits from outside sources due to errors by a non-master?
 - Injury or death to individual or others?
- What are the consequences of denying a master (high performer) the opportunity to pass?
 - Denial of deserved advancement?
 - Demoralization?
 - Lost talent for the Coast Guard?
 - Legal action if testing can be proven to be unfair?

The level of comfort or tolerance of each of the above questions will help decide how long a test should be. If the Coast Guard is willing to accept a higher risk, then the test can be shorter. If not, the test must be longer, and therefore, more reliable and valid.

Test Construction (continued)

Time Allotted for Test

Studies indicate that the average time that should be allotted for each multiple choice test item ranges from 42 seconds to 90 seconds. Time spent on a test question appears to correlate with the difficulty or complexity of the test and each item within the test. This correlation is used to determine how much time should be allotted to a given test. If a test designer decides to generate a 100 question test of moderate difficulty, the time allotted for the test should be about 1.83 hours:

- 66 seconds per test item assuming the midpoint between 42 and 90
- $66 \times 100 = 6600$ seconds
- $6600/60 = 110$ minutes
- $110 \text{ minutes}/60 = 1.83$ hours

Using a simple Likert type scale to judge time may simplify the task:

Test Difficulty	Easy	Moderately Easy	Moderate	Moderately Difficult	Difficult
Time per test item	42 seconds	54 seconds	66 seconds	78 seconds	90 seconds
50 item test time	.58 hours (35m)	.75 hours (45m)	.92 hours (55m)	1.08 hours (1h 05m)	1.25 hours (1h 15m)
100 item test time	1.16 hours (1h 10m)	1.50 hours (1h 30m)	1.83 hours (1h 50m)	2.17 hours (2h 10m)	2.50 hours (2h 30m)

Table 2-1

As a practical example from the Coast Guard electronic testing system, analysis of the Marine Science Technician Rating Advancement tests shows that on a 50 question test, with references available, the average time a test taker spends on a test item is 1.5 minutes. As actual test data is gathered by the use of electronic testing, the time allotted for the test can be refined.

Test Construction (continued)

Assigning Difficulty

Test item, and therefore test difficulty, can be estimated by using the modified Angoff method to set a cut/passing score. The modified Angoff method assigns a difficulty level to each test item. This method is explained in detail in Chapter 3 of this SOP and is required to be used for all high stakes criterion-referenced tests used in the Coast Guard. It is also recommended to be used for all medium stakes tests.

Difficulty as it relates to Angoff score for a 4-alternative multiple choice test item is:

- .25 - .49 = Hard
- .50 - .74 = Moderate
- .75 - .95 = Easy

Consider Objective Domain Size

The length of a test is also dictated by the domain size of the objectives covered on the test. If the domain size calculation requires a 300 item test and it is considered to be a difficult test, the test time required would be 7.5 hours. Using this example, it becomes obvious that test designers have to strike a delicate balance using all available information to make the best use of available time and resources while still meeting the objective of the test.

Determining Number of Test Items

The number of test items required is directly influenced by the course objectives or RPQs the test is designed to assess. Generally, the smaller the domain (size) of content derived from the objectives/RPQs, the fewer test items required to assess the objective. An example of an objective with a small content domain is:

- Without assistance, LIST the steps required to actuate a personal EPIRB, without error.

It would be difficult to write more than one or two test items to assess the objective. Most objectives, when properly constructed and written require more than one item to assess them properly. These items must be parallel to ensure adequate coverage.

Consider this objective:

- Given a schematic of the main control board and all subsystems for a National Security Cutter and access to all shipboard and manufacturers' technical publications, DIAGNOSE the cause of voltage loss to the CHT system without error.

This objective is overwhelming and therefore has a potentially large content domain from which to develop test items. The more ambiguous the objective, the harder it is to determine test length.

Test Construction (continued)

Homogeneity of Objectives

Another consideration is how closely the objectives being tested are related. This is known as “*homogeneity*.” Consider these two examples:

- Given a multimeter, an ammeter, an electrical wiring schematic, personal protection equipment, and references, TROUBLESHOOT common faults in Alternating Current (AC) electrical circuits in accordance with references.
- Given a multimeter, an ammeter, an electrical wiring schematic, personal protection equipment, and references, TROUBLESHOOT common faults in Direct Current (DC) electrical circuits in accordance with references.

These two objectives are closely related in that the content they cover is very similar. As a result, test takers will predictably perform the same way on separate test items that are written to these objectives. In technical testing terms, the objectives as well as the test items are *positively correlated*. Because these two objectives are so closely related, fewer test items are necessary to assess each objective. Only the differences in the objectives will require separate items. In the above example, any different procedures, measurements, etc. required between AC and DC circuits should be tested separately. In this case, the test length will be reduced.

If it is difficult to determine how homogeneous the objectives are, it may become necessary to include test items that fully cover each objective and, after evaluation of statistical data, reduce the coverage of parts that are closely related or identical.

Testing of EO's Required

CG TRASYS SOP Volume 3 (Evaluations) requires that all enabling objectives (EO) be tested because the EOs are the building blocks of the final performance or Terminal Performance Objectives (TPO) SOP3 states, in part:

“Enabling Objectives (EO) are the building blocks of desired performance (i.e. TPOs). EO may be tested in a variety of formats: verbal response, observed behavior, pen and paper tests (e.g. quizzes), etc. Typically, EO tests will not have the validity and reliability to make a final judgment about a student’s performance. They should be used as a progress check to redirect a student’s learning. Level 2 evaluations should test TPOs and EOs directly and avoid ‘nice to know’ additions.”

If the EO is not assessed, there is no way of confirming the candidate has completed or has knowledge of all of the required steps to support the required performance. The EOs can be assessed in a variety of formats including: performance tests, criterion-referenced checklists, verbal response, pen and paper, computer based knowledge or application of knowledge tests, etc.

Test Construction (continued)

ERATS

The Enlisted Rating Advancement Training System (ERATS) has reshaped the way the USCG prepares the enlisted workforce for advancement. A detailed explanation of ERATS can be found in CG TRASYS SOP Volume 9 (ERATS). Briefly, ERATS is designed to support the performance of assigned enlisted personnel in preparing for advancement or change in rating. To accomplish this, ERATS functions to:

- Define performance requirements
- Provide training and performance support
- Assess performance

Testing of RPQs

Enlisted performance requirements are defined by the Rating Performance Qualifications (RPQ) which are derived from the Occupational Analysis (OA). Each RPQ is supported by performance tasks (TPO) that have task steps (EO) associated with them, i.e. the performer must complete each step to satisfactorily complete each performance task and complete each performance task to satisfy the requirements of the RPQ. To verify the satisfactory completion of each RPQ, each of the task steps must be assessed by some method. Much the same as assessing EOs, this can be accomplished in a variety of ways including: performance tests, criterion-referenced checklists, verbal response, pen and paper, computer-based knowledge or application of knowledge tests, etc. It is important to not confuse the supporting sub-steps with the actual task step or EO. An oversimplified example might be:

- RPQ (TPO) – CONSTRUCT a non-load bearing 2x4 framed wall...
 - Task Step (EO) – DRIVE a common nail...
 - Task Sub-Step – DETERMINE correct nail to use
 - Task Sub-Step – SELECT correct type of hammer

TPOs, EOs, and Test Items

In a resident training "A" school environment, each RPQ and associated performance tasks and steps at the E-4 level are translated into TPOs and EOs and, possibly, step level lesson objectives to facilitate instruction of the final required performance. In the resident "C" school environment, the performance tasks are identified by the formal analysis (FEA or JTA) from which the course designer develops TPOs and EOs. Test items shall be developed at the EO or step level.

Test Construction (continued)

Role of Objectives in Testing

The creation of good instructional objectives is not only essential in the creation of good instruction but is critical in the creation of sound, valid,

reliable, and defensible test items. The components of well written objectives translate directly to well written test items. Instructional objectives serve three important purposes for a test item developer:

- Test items written to objectives ensure that a test covers those learner outcomes that are deemed important enough to be included in course material. Matching test items to appropriate objectives ensures all essential content is assessed.
- Sound objectives increase the accuracy with which cognitive processes can be assessed. A well written objective is essentially a blueprint for the creation of test items that will assess specific requirements or competencies required by the objective. These objectives make it much easier to create parallel test items so the requirements of the objective can be tested in several ways to ensure complete understanding of the required performances or outcomes.
- Design of objectives must consider the size of the domain to be covered, the homogeneity of the objectives, and the objectives of the tests designed to assess the domain or area of required performance/knowledge. These are important factors in determining how many test items will be required to assess the participants.

Research on Test Length

Research has shown that the accuracy of assessments or tests is a direct function of the number of test items per objective with accuracy improving as the number increases. It has also been found that the improvement in accuracy when testing less critical objectives tends to level off at 4 to 6 test items per objective. As the criticality of the objectives increases, the number of test items to ensure adequate coverage should also increase. Objectives related to behaviors critical to safety, health, legal requirements, etc. should be tested with more than six test items and possibly several times especially if the content domain of the critical objectives is large (Shrock, S. & Coscarelli, W. 2007).

Test Construction (continued)

Number of Test Items per Objective

Table 2-2, adapted from Shrock, S. & Coscarelli, W. 2007, shall be used as a first estimate guideline for determining the number of test items per objective or RPQ task step to be included on a test.

If	--	And	And	Then
The performance objective is:	Critical to safety, life, limb, legal requirements, etc.	From a large objective domain	Unrelated	10-20
			Related	10
		From a small objective domain	Unrelated	5-10
			Related	5
	NOT critical to safety, life, limb, legal requirements, etc.	From a large objective domain	Unrelated	6
			Related	4
		From a small objective domain	Unrelated	2
			Related	1

Table 2-2

After statistical results of the test have been reviewed, the number of test items per objective may require refinement. If you find that reliability of a particular objective is less than expected or desired, additional test items may be needed or the existing test items may require revision. It is also important to strike a good balance between the number of test items and the allotted time in the curriculum to complete each test.

Minimum Number of Test Items per RPQ/Objective

It is important to maintain a robust database of test items to allow for variation and randomization of tests. In the case of the Servicewide Examination and the Rating Advancement Test, there must be sufficient test items available to generate parallel examinations which equally test the same RPQs or objectives at the same level of difficulty. Keep in mind that the more an individual test item is exposed, the higher the chance there is for compromise. For this reason, it is highly recommended that a minimum of three to five test items per RPQ step or Enabling Objective be maintained in the active database. Any fewer than the minimum increases the chances of overexposure and duplicated items on parallel test forms.

Test Construction (continued)

Balancing Number of Test Items with Allotted Time

As tests are developed, it is important to strike a balance between the number of test items required to adequately test the desired domain of the objective or RPQ and the time allotted to administer the test. Table 2-2 should be used as the initial determination of number of test items required. Once the initial number is determined, the table 2-1 should be consulted to establish allowed test time.

Defensibility of Tests

Validation Required

Once a test or assessment is designed and developed, it must be administered and analysis must be performed. All tests and assessments utilized in the Coast Guard must be valid, reliable, and defensible. The first steps towards this are to ensure all test items can be related to an authoritative reference and actually test the objectives that they are designed to test. Next is to have a plan to follow. Test plans strengthen the defensibility of tests and are highly recommended for all high stakes tests.

Elements Required

Derived from the *Standards for Educational and Psychological Testing* (AERA, APA, NCE, 1999), the U.S. Air Force, U.S. Navy, and U.S. Coast Guard representatives developed best practices and recommended elements to maintain defensibility of tests. Based on the elements most applicable to military enlisted advancement testing, and using the 12 Steps for Effective Test Development from the *Handbook for Test Development* (Downing & Haladyna, 2006)), the recommendations include the following: (Standards referenced are inside parenthesis):

1. Overall plan (1.1, 3.2, 3.9)
2. Content definition (1.6, 3.2, 3.11)
3. Test specification (1.6, 3.2, 3.3, 3.4, 3.11)
4. Item development (3.6, 3.7, 3.17, 7.2, 13.18)
5. Test design and assembly (3.7, 3.8)
6. Test production
7. Test administration (3.18, 3.19, 3.20, 3.21)
8. Scoring test responses (3.6, 3.22)
9. Passing scores (4.10, 4.11, 4.19, 4.20, 4.21)
10. Reporting test results (8.13, 11.6, 11.12, 11.15, 13.19, 15.10, 15.11)
11. Item banking (6.4)
12. Test technical report (3.1, 6.5)

Test Plan

The same elements that are required to maintain defensibility (listed above) provide the structure for the test plan. A sample test plan and a standard template are provided as Appendix D to this SOP. Although the information in the test plan appears somewhat lengthy, it consists of many “boilerplate” items that will only require minor editing. If maintained, the test plan shall be kept electronically in a secure location for three years after the test is last administered and then the test plan may be disposed of.

Chapter 3

DETERMINING TEST CUT/PASSING SCORES

Overview of the Modified Angoff Method

Introduction

To be legally defensible and meet the Standards for Educational and Psychological Testing, cut scores for tests cannot be arbitrarily determined. The U.S. Coast Guard has chosen the modified Angoff Method to set cut scores. The procedures described in this chapter are mandatory when setting cut/passing score for all high stakes criterion-referenced tests developed for use within the US Coast Guard training system and recommended for medium stakes tests.

The American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) jointly developed the Standards for Educational and Psychological Testing. In addition to providing testing standards, the *Standards* also addresses professional and technical issues of test development and use and presents measurement trends affecting validity of tests.

William H. Angoff

William H. Angoff, an expert on measurements used in testing and scoring, co-authored the *Educational Measurements* book in 1971, where he wrote in a footnote:

“...keeping the hypothetical ‘minimally acceptable person’ in mind, one could go through the test item by item and decide whether such a person could answer correctly each item under consideration.”

“... ask each judge to state the probability that the ‘minimally acceptable person’ would answer each item correctly.”

This footnote was the origin of the Angoff Method, a standard-setting process designed to support the defensibility of a cut score.

2014 Decision by USCG Legal Staff

If in the judgment and experience of the test developer the modified Angoff method is not justified, then the TRACEN Commanding Officer may make a written request to FORCECOM (FC-T) to waive any or all of these test development procedures. Only with a written waiver from FORCECOM (FC-T) may a TRACEN deviate from these standards.

Upon collection of an adequate sample of actual test results, the modified Angoff method cut-off scores may be revised to reflect actual responses and difficulties for each test item. In this process, cut-off scores are calculated using the actual data through the modified Angoff method algorithms.

Applying the Modified Angoff Method

Validity / Reliability

For a test to be legally defensible, two standards must be met:

1. **Validity:** The test must measure what the students are expected to know. This is accomplished by writing test questions that align with the objectives.
2. **Reliability:** The test must produce consistent results time after time. That is, the test should produce the same score if administered to the same students again and again.

To ensure validity and reliability, it is best to conduct pilot assessments and statistical analyses of test items prior to applying the modified Angoff method.

The Modified Angoff Method

The Modified Angoff Method is a process that determines how often a minimally qualified performer would answer a test item correctly. A panel of experts is chosen to review test items and estimate the probability that a minimally qualified performer would answer the items correctly. The estimates for each test item are averaged, and those averages are used to determine the cut score.

While the Modified Angoff Method can be used for performance tests, the information provided in this section applies to written/computer-delivered criterion-referenced assessments only.

Reviewing these types of assessments using the Modified Angoff Method is a dedicated project. Raters must be chosen, a site must be available, and time must be afforded. There are five steps involved:

1. Select and gather the raters.
2. Identify the “minimally qualified performer.”
3. Rate the items.
4. Review the ratings.
5. Determine the cut score.

For step-by-step procedures and examples of spreadsheets used in the process, refer to Appendix C.

Adding New/Revised Test Items

The addition of new or revised test items requires that an Angoff rating panel be convened to rate the new items and recalculate the overall test cut score prior to adding the items to the “live” database. If a new score is not calculated and tests are generated randomly, fairness and defensibility of the test instrument is lost.

CHAPTER 4

SECURITY, COMPROMISE, AND DESTRUCTION

Background

Introduction

Everyone involved in the development and administration of testing material has the responsibility of safeguarding material in order to prevent compromise. Revision of a compromised test is expensive and time-consuming. A compromise could delay promotions, enable less than qualified personal to fill vital positions, etc. Therefore, general guidelines have been included here to prevent compromise. Detailed instructions can be found in Physical Security and Force Protection Program, COMDTINST M5530.1 (series); Classified Information Management Program Manual, COMDTINST M5510.23 (series); Coast Guard Freedom of Information and Privacy Acts Manual, COMDTINST M5260.3 (series); and Information Assurance Manual, COMDTINST M5500.13 (series) and DHS Management Directive MD-11042.

Safeguarding Testing Materials

All unclassified testing material is considered sensitive and is afforded protection as such. All classified material will be afforded the protections of its classification. Testing material includes the test plan, draft and camera-ready copies, answer keys, work sheets and notes, statistical data, and any hard copies of test questions. All elements of computer technology that contain the above listed sensitive material must also be safeguarded.

Due to the nature of the contents, impact of unauthorized disclosure, and Personal Identifiable Information (PII), the Rating Advancement Test and the Servicewide Examination are considered to be Controlled Unclassified Information (CUI) (unless the material is actually classified). As such, all RATs and SWEs shall be considered and marked as For Official Use Only (FOUO). Additional details are contained in Chapter 6 of this SOP.

Background (continued)

Terminology

Terms associated with security and compromises of sensitive or classified information are listed in table 4-1.

Term	Definition
Access	The ability and opportunity to obtain knowledge or possession of classified information. An individual is not considered to have access to classified information merely by being in a place where such information is kept, that is, provided the security measures which are in effect prevent the individual from gaining knowledge or possession of such classified information.
Classified material	Any matter, document, product, or substance on or in which classified information is recorded.
Compromise	The disclosure of sensitive or classified information to persons not authorized access.
Confidential	The designation which shall be applied to information or material the unauthorized disclosure of which could reasonably be expected to cause damage to national security.
Controlled Area	<p>A controlled area is the least secure type of restricted area. It may also serve as a buffer zone for exclusion and limited areas, thus providing administrative control and protection against sabotage, disruption, or potentially threatening acts.</p> <p>Uncontrolled movement may or may not permit access to the security interest or asset.</p>

Table 4-1

Background (continued)

Definitions

Term	Definition
For Official Use Only (FOUO)	Unclassified information of a sensitive nature, not otherwise categorized by statute or regulation, the unauthorized disclosure of which could adversely impact a person's privacy or welfare, the conduct of Federal programs, or other programs or operations essential to the national interest.
Password	A protected word or string of characters that identifies or authenticates a user for access to a specific resource such as a data set (file) or record.
Physical Security	Internal security concerned with the physical measures designed to prevent unauthorized access to equipment, facilities, material, and documents and to safeguard them against espionage, sabotage, damage, theft, or other acts which would in some degree lessen the ability of the command to perform its mission or would affect overall national security interests.
Restricted Area	Any area to which access is subject to special restrictions or controls for reasons of security or safeguarding of property or material. This term is a legal designation. Specific administrative terms are used to designate various security areas/levels.
Sensitive Material	Material which requires a high degree of protection and control due to regulatory requirements.

Security of Test Materials

Personnel Designation

All personnel having access to high stakes testing material (e.g. RAT, SWE, EPME, etc.) shall be designated in writing by the unit's commanding officer. These personnel shall be required to execute a non-disclosure statement (see Appendix E) in which they agree to not discuss or divulge in any manner, the contents or design/weighting of the testing material. The non-disclosure remains in force during their assignment and for a period of not less than three years after assignment end.

Personnel having access to low and medium stakes testing material shall be authorized as directed by local command policy.

Work Site Designation

To reduce the risk of compromise, all spaces where production, review or access to high stakes or classified test material occurs shall be designated as a "**Restricted Area – Controlled**" and posted as a "Restricted Area" in accordance with the Physical Security and Force Protection manual, COMDTINST M5530.1 (series).

Work areas where production of low and medium stakes testing material occurs shall be considered non-restricted areas but should have limited access with only authorized personnel permitted in the vicinity when test production or viewing is in progress.

Posting of Classified Restricted Areas

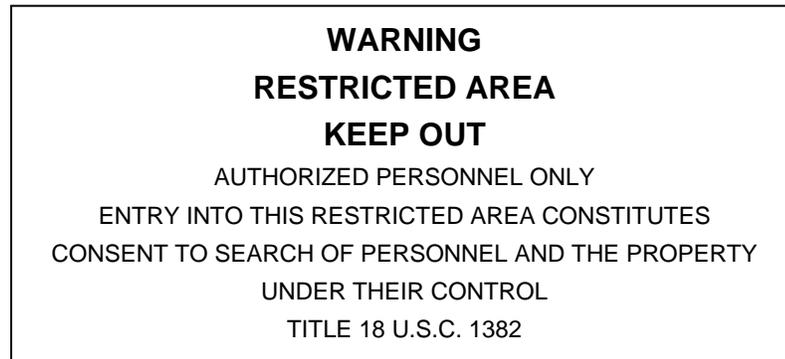
Restricted areas that are also classified spaces, shall be posted at all external points of entry with signs approximately two feet by two feet in size (or a size deemed appropriate by the cognizant SECMGR) with appropriate lettering. Signs shall read:

<p style="text-align: center;">WARNING</p> <p style="text-align: center;">RESTRICTED AREA</p> <p>THIS AREA HAS BEEN DECLARED A RESTRICTED AREA IN ACCORDANCE WITH TITLE 18 U.S.C. 1382. UNAUTHORIZED ENTRY IS PROHIBITED. ALL PERSONS ENTERING HEREON ARE SUBJECT TO SEARCH. PHOTOGRAPHING OR MAKING NOTES, MAPS, DRAWINGS OR GRAPHIC REPRESENTATIONS OF THIS AREA OR ITS ACTIVITIES IS PROHIBITED UNLESS SPECIFICALLY AUTHORIZED BY THE COMMANDING OFFICER. ANY SUCH MATERIAL FOUND IN THE POSSESSION OF UNAUTHORIZED PERSONS SHALL BE CONFISCATED.</p>
--

Security of Test Materials (continued)

Posting of Non-classified Restricted Areas

Other restricted areas shall be posted at all external points of entry with signs approximately two feet by two feet in size with appropriate lettering. Adjustments to the size of the sign to facilitate placement on doors is authorized, as long as the required wording remains unchanged and legible. Size should not be adjusted any smaller than is minimally necessary and the wording must remain clearly readable from a distance. Signs shall read:



Access Control

Admission into a “restricted area- controlled” is granted only to persons whose duties require access and who have been granted appropriate authorization. Individuals without adequate identification as determined by the commanding officer shall be logged in/out.

A personnel identification and control system, including an access list posted inside the area shall be maintained. During normal duty hours, use of an entry/departure log is suggested, but not required. After normal duty hours, all personnel accessing the controlled area shall be logged in/out.

Admission into a restricted area that contains classified material requires specific control procedures set forth by the Physical Security and Force Protection manual, COMDTINST M5530.1 (series).

Visitor Control

Visitors to test production areas should be kept to a minimum. If, at any time, personnel visit workspaces where testing material is being processed, the testing material should immediately be put out of sight and the visit terminated until a more opportune time. A visitor control log is not required but should be maintained at all restricted areas.

Security of Test Materials (continued)

Key and Lock Control System

A key and lock control system shall be established for all restricted areas. A key and lock control system supplements other security measures used to control access and are essential for the safeguarding of high stakes testing material.

All restricted areas shall be equipped with cipher locks. Under no circumstances shall a key to or the combination to any cipher lock be in possession of personnel who do not work in or have written authorization to enter the specified area.

If entrance is required after normal working hours, a staff member must be recalled to open the space. The officer of the day (OOD) shall have letters designating staff for recall for entry in each restricted area.

Specific guidelines to establish the system are outlined in the Physical Security and Force Protection Program, COMDTINST M5530.1 (series).

Storage of Test Materials

All unclassified testing material is considered to be sensitive material and should be afforded a high degree of security. Testing material includes camera-ready copies, work sheets and notes, statistical data, any hard copies of test questions and any electronic/magnetic media containing test materials.

Test materials shall never be left unattended. Typically, low to moderate stakes test material may be stored in a locked file cabinet or desk to prevent unauthorized access. All high stakes testing materials, including the Rating Advancement Test and the Servicewide Examination, are considered to be Controlled Unclassified Information (CUI) (unless the material is actually classified). As such, shall be considered and marked as For Official Use Only (FOUO) and should be stored in a GSA approved security container with a combination lock.

All elements of computer technology containing high stakes testing materials shall be encrypted and should be stored in a GSA approved security container with a combination lock. In no case shall copies of high stakes tests be stored on the local drive of a networked computer. The servers and repositories containing test material within the Coast Guard enterprise testing software application are considered to be secure.

Security of Test Materials (continued)

Safe and Cipher Lock Combinations

Safe and cipher lock combinations and the Security Container Check Sheet (SF-702) shall be maintained in accordance with Classified Information Management Program, COMDTINST M5510.23 (series). Access to testing material must be limited to authorized persons (those with the proper clearance and with a need to know). The local Test Development Manager shall maintain the sealed copy of the SF-702 for all cipher locks and security containers containing high stakes testing materials. This copy shall be stored in a GSA approved security container accessible only to the local Test Development Manager and alternate.

Changing Combinations

Combinations to cipher locks and security containers used to store testing material shall only be changed by the custodian. Combinations shall be changed under any of the following circumstances:

- Upon receipt of the container/lock
- Whenever the custodian is transferred, discharged, or reassigned
- When the combination or record of the combination has been compromised, or when the security container has been discovered unlocked and unattended
- At least annually (this applies to classified material containers only)

When selecting a combination, DO NOT use:

- Multiples of five or simple ascending or descending arithmetical series
- Personal data such as a birth date or SSN

Note: Combinations shall NOT be carried in wallets, purses or otherwise on the person or hidden within an office space.

Security of Test Materials (continued)

Security Container Precautions

The following precautions shall be followed to ensure that adequate security is being provided for testing material:

- Security containers shall be kept locked when not under the direct observation of the custodian or other authorized persons.
- Reversible "Closed/Open" signs shall be used as additional reminders on security containers.
- For a safe containing classified material, a Security Container Check Sheet (SF-702) shall be affixed to each container for the purpose of checking each time the container is opened or closed.
- Only testing material should be stored in the container. This includes paper and electronic/magnetic media devices.
- Testing material not being immediately processed shall be properly stored.

Note: Additional information can be found in the Classified Information Management Program, COMDTINST M5510.23 (series)

Security of Test Materials (continued)

Local Movement of Testing Materials

When testing material is being moved within workspaces or around the unit, the material must be strictly controlled. Each individual must ensure that access is limited to authorized personnel only. The following steps must be taken to safeguard the material:

A cover sheet must be used when handling testing material within workspaces. For SWEs and RATs, in addition to the required DHS FOUO cover sheet a locally designed file folder or other container marked with minimum 1/4" letters:

SENSITIVE MATERIALS
TO BE OPENED ONLY BY A
TESTING MATERIALS OFFICER

For classified material, the appropriate cover sheet or label shall be affixed to the inner folder. For movement of testing material outside a building, insert the testing material (along with the cover sheet/folder) into another container (a sealed envelope or a briefcase).

- Turnover of testing material must be from hand to hand. At no time shall testing material be left in an office when persons authorized to receive it are not present.
 - Testing material shall not be viewed, studied, displayed, or worked on except in authorized spaces.
-

Security of Test Materials (continued)

Traveling With Testing Materials

Hand carrying test material when you are traveling should be kept to an absolute minimum. However, if the situation arises where it cannot be avoided, the following safeguards must be in place:

- Testing material must be kept in the physical possession of the individual at all times unless proper storage at a U.S. Government activity is available.
- Testing material shall not be viewed, studied, displayed, or worked on while in public conveyances or places.
- Individuals responsible for hand-carrying classified testing material in travel status must be authorized in writing by the commanding officer. A clearance verification letter or message may serve as written authorization. However, one-time authorization letters (courier letters) shall be used for individuals' hand-carrying classified testing material in a travel status. These individuals must also be briefed on their duties and responsibilities in accordance with the Classified Information Management Program, COMDTINST M5510.23 (series).

Note: Additional requirements for a courier letter can be found in the Classified Information Management Program, COMDTINST M5510.23 (series).

Packaging of Testing Materials

All testing material, whether sensitive or classified, shall be mailed/shipped double-wrapped with the inside envelope containing the following instructions on both sides in at least 1/4-inch letters:

SENSITIVE MATERIALS
TO BE OPENED ONLY BY A
TESTING MATERIALS OFFICER

The office and person designated to receive testing material shall be identified on the inner envelope only.

Security of Test Materials (continued)

Mailing or Shipping of Testing Materials

When sensitive high stakes testing material is mailed between the training source (TS) and CGI, PPC (adv), or other authorized activity, it must be accounted for by signature using authorized overnight or two-day delivery service (FedEx). Classified testing material must be mailed via registered mail. DO NOT use certified mail. A mail logbook shall be used to track testing material that is mailed or shipped.

Electronic Transmission of Sensitive Testing Material

Use of USCG e-mail to send sensitive testing material shall be in accordance with the following procedures set forth in the Security and Information Assurance (SIA) Manual, COMDTINST M5500.13 (series):

- E-mails containing PII/SPII sent to .MIL and DHS.GOV domains shall be encrypted and digitally signed.
- Sensitive government information that is sent to authorized users who reside outside the DOD/DHS domain on a commercial or non-government domain to include but not limited to (.com/.net/.edu/.org, etc.) shall have all sensitive information sent as a protected attachment to the e-mail. No sensitive information shall reside in the e-mail body. The attachment shall be encrypted, (using WINZIP (256 bit Advanced Encryption System (AES)) or within a password protected WORD/ADOBE document. The password to the encrypted file or protected attachment shall be provided to the receiver of the original e-mail via separate correspondence. (*The USCG Academy (@uscga.edu) network **is not** part of the USCG.MIL domain; therefore the above requirements apply*).
- Passwords of encrypted file(s) shall be sent via a separate e-mail with a different subject line. The password shall meet the strong password criteria.
- Transmission of SBU government information to personal e-mail accounts is **prohibited**.
- The CG Outlook Webmail Application (OWA) is an authorized CAC/ALAC enabled capability to access CG e-mail from the Internet
- Avoid selecting “Replying to all” on any e-mail with SBU information. Users shall review and remove any e-mail recipient (including distribution groups) that do not need to know the response (e.g., Command decision) and/or information to perform their duties.

Use of both the Delivery Receipt and Read Receipt features of Outlook is required when transmitting sensitive test material via e-mail.

Security of Test Materials (continued)

Mailing or Shipping of Classified Testing Materials

Classified testing material shall be handled, mailed/shipped, and disposed of in accordance with the Classified Information Management Program, COMDTINST M5510.23 (series) and Physical Security and Force Protection Program, COMDTINST M5530.1 (series).

Telephone and Messaging Security

The discussion of sensitive testing material over the telephone shall be avoided. The discussion of classified matters on the telephone is **STRICTLY FORBIDDEN**.

Use of instant messaging features such as Microsoft Office Communicator and unsecured e-mail to discuss content of testing material is prohibited.

Note: All government and non-government electronic devices are subject to monitoring and recording at all times.

Disposal and Destruction of Testing Material

All work materials, (i.e., notes, research papers, drafts, etc.) used for development of high stakes test items (RAT, SWE, etc.) must be disposed of by authorized individuals through shredding or burning in accordance with the Freedom of Information Privacy Acts Manual, COMDTINST M5260.3 (series) and the Classified Information Management Program, COMDTINST M5510.23 (series). When shredding, a crosscut shredder must be utilized at a minimum for all unclassified testing materials.

Test items that are stored in an electronic data base shall be deleted 3 years after they are retired from use. This action is restricted to local e-testing managers and system managers to ensure items are not deleted inadvertently.

Although not addressed specifically in the Information and Life Cycle Management Manual, COMDTINST M5212.12 (series), all previously administered high stakes tests shall be destroyed after 3 years. All other test material (not results/grades) may be destroyed when the test is superseded or obsolete.

Security of Test Materials (continued)

Reproducing Testing Materials

High stakes testing material shall not be reproduced unless approved by either the RKM for the rating or higher authority. Once reproduced, the new copy shall be accounted for in the Test Material Control Log for the appropriate rating. Personnel reproducing test material shall exercise care to prevent the testing material from being compromised. They shall be aware of and alert others concerned to the specific security hazards listed in chapter 7 of Classified Information Management Program, COMDTINST M5510.23 (series).

Caution must be exercised when reproducing any testing material on a copy machine or a printer. Most modern copy machines and printers contain internal memory storage devices that retain electronic versions of the material until erased or overwritten.

Passwords

All passwords used in association with sensitive testing materials shall be formatted as a strong password with a minimum of eight characters (14 – 20 are recommended) composed of at least one upper case, one lower case, one number, one special character (avoid the following: & , “ / \ ‘ | < > : ; # £) and shall:

- Not be the same as the previously used 8 passwords for the application/e-mail, etc.
 - Be changed every 90 days, or at the completion of any large/major unit event that allowed access
 - Not contain any dictionary word in any language
 - Not contain any proper noun or the name of any person, pet, child, or fictional character
 - Not contain any EMPLID number, Social Security number, birth date, phone number, or any information that could be readily guessed about the creator of the password
-

Security of Test Materials (continued)

Labeling of Removable Media

All removable media (CR-ROM) shall be labeled in accordance with the sensitivity or classification of the data stored on them. The following standard forms may be used:

- Unclassified: SF-710
- Confidential : SF-708
- Secret: SF-707

Note: Standard forms can be ordered through the Coast Guard acquisition system.

Other types of removable media such as encrypted hard drives shall be labeled appropriately (e.g. UNCLASSIFIED – FOUO).

Compromise

Compromise Defined

A compromise is the disclosure of information to a person or persons who are not authorized access to that information. The unauthorized disclosure may have occurred unknowingly, willfully or through negligence. Compromise is confirmed when conclusive evidence exists that information has been disclosed to unauthorized person(s).

A compromise may take several forms including but not limited to:

- Tests, answer keys or completed participant answer sheets left unattended
- Copies or transcriptions of actual test items made by photo, text message, web/blog postings, notes, etc.
- Testing material lost or unaccounted for by the examination board or ESO
- Review of high stakes test items with candidate by ESO or other member after test administration
- Discussion of actual test items by the candidate upon completion of the test
- Safe or other secure container for storage of testing materials left open and unattended
- Non-receipt of test material that was mailed, shipped or electronically transmitted
- Testing material sent to an unknown networked printer
- Candidate utilizing e-mail, instant messaging, text messaging, search engines or other communications during test administration
- Loss or inability to account for testing materials during production

Note: The posting or sharing of “study guides” on Web sites or paper or the formation of study groups is not considered a compromise as long as actual test items that are in production do not appear in the material or discussion. Test developers are encouraged to monitor these activities on a routine basis and report suspected compromises in accordance with the guidance in this chapter.

Compromise (continued)

Initial Action Required

Any civilian employee, military personnel, or other person associated with the Coast Guard, having knowledge of the loss, unauthorized disclosure, or possible compromise of a high stakes test (RAT, SWE, EPME, etc.) or of an infraction of security regulations should immediately advise their command security officer. Once an incident is disclosed, commands will ensure that the matter is reported immediately in accordance with the procedures set forth in this chapter.

Compromise of SWE

Once an incident is disclosed, commands will ensure that the matter is reported immediately in accordance with the procedures set forth in the CIM 1000.2, Article 3, B.4.d and this chapter.

In any case where the loss or compromise of a Servicewide examination is discovered or circumstances indicate that loss or compromise is possible, a full report shall be immediately forwarded to Commanding Officer CG PPC (adv) with a copy to Commander CG PSC-EPM-1.

This initial report may be sent via message (preferred), email or memo and should include as a minimum:

- The title of exam(s) involved
- A brief summary of the circumstances known
- A preliminary evaluation as to the probability of compromise (Low, Medium, High)
- POCs

If all facts are not readily available, the commanding officer should convene an investigation in accordance with reference (j), Administrative Investigations Manual, COMDTINST M5830.1 (series). The record of proceedings in the investigation shall be forwarded in accordance with reference (j), Administrative Investigations Manual, COMDTINST M5830.1 (series), with advance copies submitted to Commander CG PSC-EPM-1, Commanding Officer CG PPC (adv) and the Coast Guard Investigative Service in amplification of the report required above.

Compromise (continued)

CGMS Message Format for Reporting SWE Compromise

The following format shall be utilized for notification via the Coast Guard Message System:

R_____Z MMYYY
 FM (UNIT PLAD)
 TO COGARD PPC TOPEKA KS//ADV
 INFO COGARD INVSER HQ WASHINGTON DC
 COGARD FORCECOM WASHINGTON DC//T//
 COGARD PSC ARLINGTON VA//EPM-1//
 COGARD TRACEN YORKTOWN VA//T//
 COGARDTRACEN PETALUMA CA//T//
 COGARD AVTECHTRACEN ELIZABETH CITY NC//T//
 COMCOGARD MLE ACADEMY CHARLESTON SC//T//
 BT
 UNCLAS//N01418//
 SUBJ: POSSIBLE COMPROMISE OF (RATING PAYGRADE)
 SERVICEWIDE EXAMINATION
 A. ART 3, B.4.D OF COMDTINST M1000.2 (SERIES)
 B. COGARDTRASYS SOP VOL-10 CHAPTER 4
 1. TITLE OF TESTING INSTRUMENT(S) (e.g. BAI SERIES 84)
 2. PROBABILITY OF COMPROMISE (HIGH, MEDIUM, LOW)
 3. (BRIEF SUMMARY OF CIRCUMSTANCES SURROUNDING
 SUSPECTED COMPROMISE)
 4. (BRIEF DESCRIPTION OF ACTION TAKEN BY UNIT)
 5. POC IS (RATE, NAME, TEL, E-MAIL)
 BT

Reporting SWE Compromise via Email or Memo

Although initial notification of a possible compromise of a SWE by CGMS is preferred, notification by email or memo is acceptable for low probability of compromise. If submission via e-mail or memo is chosen, include the above information in e-mail or memo format to the same recipients.

Compromise (continued)

Compromise of RAT, EPME, or AQE Tests

All EOCT/RAT Potential Compromise situations should be processed in accordance with CG Institute SOP Volume II. Once an incident is disclosed, commands will ensure that the matter is reported immediately in accordance with the procedures set forth in the CG Institute Note 1000.2 Vol. 2, and this chapter.

A compromise is the disclosure of EOCT/RAT information to a person(s) who/whom is not authorized access to that information. The unauthorized disclosure may have occurred unknowingly, willfully or through negligence. Compromise is confirmed when conclusive evidence exists that EOCT/RAT information has been disclosed to an unauthorized person(s).

Any civilian employee, military personnel, or other person associated with the Coast Guard, having knowledge of the loss, unauthorized disclosure, or possible compromise of EOCT/RAT information or of an infraction of security regulations should immediately advise their command security officer. Once an incident is disclosed, commands will ensure that the matter is reported immediately in accordance with the procedures set forth in this publication.

The steps for reporting a compromise are as follows:

- Contact the CG Institute CST Division via email, stating the circumstances of the possible compromise.
- Coast Guard Institute will determine if a potential compromise exist. If a potential compromise does exist, CG Institute will email the respective command a memo requesting that command to initiate an administrative investigation in accordance with the Administrative Investigation Manual, COMDINST M5830.1 (series). Investigation shall be completed within 15 business days. If an investigation is not required, the CG Institute will advise what action to take via email.
- Once the Administrative Investigation is complete, the applicable Course Writer will make a recommendation to Compromise, or Not Compromise the exam to CG Institute.
- CG Institute will notify ESO network via TACCTS if a test has been declared compromised.

Authority to Suspend Use of RAT/EPME

FORCECOM (FC-TD) has delineated the authority for the Commanding Officer of the Coast Guard Institute to suspend EOCT/RAT, while an investigation for determination of a possible compromise is being conducted.

The CG Institute will notify ESOs and RKM via TACCTS, of any required action.

Compromise (continued)

Procedures Upon Confirmed Compromise of RAT, SWE or EPME

Should a compromise of a RAT, SWE, EPME, or AQE be confirmed, the training source must take immediate action to develop replacement test instruments using the following steps:

Step	Action						
1	Immediately retire all items that appeared on the compromised test(s) from the USCG test development database to prevent further use.						
2	Develop sufficient replacement test items to encompass each of the objectives/tasks/RPQs involved in the compromise. <table border="1" data-bbox="690 688 1403 1010"> <thead> <tr> <th>If...</th> <th>Then...</th> </tr> </thead> <tbody> <tr> <td>RAT, EPME, or AQE</td> <td>New items must be developed and Angoff rating performed within 30 days.</td> </tr> <tr> <td>SWE</td> <td>New items must be developed within 14 days.</td> </tr> </tbody> </table>	If...	Then...	RAT, EPME, or AQE	New items must be developed and Angoff rating performed within 30 days.	SWE	New items must be developed within 14 days.
If...	Then...						
RAT, EPME, or AQE	New items must be developed and Angoff rating performed within 30 days.						
SWE	New items must be developed within 14 days.						
3	Construct replacement test instruments and forward to appropriate test administration command for distribution (PPC (adv) or CGI (cst)).						

Compromises Involving Classified Material

In the case of lost, mishandled, unaccounted for, or possibly compromised classified material, the ESO should comply with the Coast Guard Security Manual COMDTINST M5510 (series).

Prohibited Actions Upon Transfer or Termination

Any personnel involved with the development of content for high stakes tests are prohibited from discussing or revealing, in any form, the contents or testing strategies employed for a period of three years after transfer or termination of employment.

THIS PAGE LEFT INTENTIONALLY BLANK

Chapter 5

TEST ITEM WRITING

General Principles

Background

Test item writing is often referred to as an art (Cantor, 1987). Designing and developing test items and ultimately testing instruments are not simple tasks. In the summer 1981 issue of *NCME Measurement in Education*, Robert L. Ebel and Samuel A. Livingston (Ebel & Livingston, 1981), world renowned experts on measurement in education, referred to test construction as follows:

Those who prepare tests of competence should:

- a) Be themselves outstandingly competent in the field.
- b) Be skilled in expressing ideas concisely and clearly.
- c) Be guided by professional advice on how to write effective test items.
- d) Be willing and able to take time to do the job well.

This section identifies various types of test items and provides guidance on how to properly construct defensible and concise items.

Three Methods of Test Item Construction

In addition to the item writing principles discussed in this section, test items must conform to the rules stated.

There are three basic methods used to construct test items and ultimately tests: objectives-based, topic-based, and statistically-based. Objectives-based is what is currently accepted and in use for Coast Guard testing. A topic-based design should never be used but is described below for comparative purposes. Statistically-based is the ideal test construction method for norm-referenced tests.

- Objectives-based test item construction is the method of choice for all U.S. Coast Guard criterion-referenced tests and until such time the statistically based method is feasible for all SWEs. In order to be defensible, a test item must be based upon an objective or a specific competency.
-

General Principles (continued)

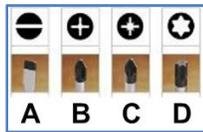
Three Methods of Test Item Construction (continued)

- When developing test items for “A” or “C” school courses, all test items shall be derived from course enabling objectives. For Rating Advancement Tests (RAT) and Servicewide Examinations (SWE), test items shall be derived from the tasks and task steps that support the Rating Performance Qualifications (RPQ), required competencies, and the references that directly support the RPQ.

An example of an objectives-based test item is:

Objective: Given a selection of previously unseen screwdrivers, the student will IDENTIFY selected screwdrivers without error.

Test item: Which of the following is a Phillips tip screwdriver?



- As a test construction method, the topic-based method is widely used because it is relatively simple and is taught in many teacher education courses. Using this method the test item writer simply generates several test items from each chapter or unit of instruction or student guides based on their own perception of importance of the topic. The problem with this method is that it is imprecise, does not allow for criterion-referenced statistical interpretation and, if used for norm-referenced testing, does not provide widespread test scores for rank ordering.
 - Topic-based item construction is not defensible and shall not be used when designing tests for use within the U.S. Coast Guard.
 - The statistically-based method is best used to construct norm-referenced tests. Test items are constructed based upon established criteria. Once the items are used as non-scoring items on norm-referenced tests to garner sufficient statistical data, a test is constructed by carefully choosing items that will produce the desired separation for rank-ordering test takers.
 - Ideally, this is how the U.S. Coast Guard Servicewide Examinations (SWE) should be constructed. Until such time as the Coast Guard has the infrastructure in place to acquire reliable SWE test item statistical data, use of this method is deferred.
-

General Principles (continued)

Definitions

The following basic definitions apply to test item writing:

Stem: The part of a test item which asks a direct question, gives a command, or is an incomplete sentence and provides a blank for the test taker to fill in missing information.

Correct Response(s): The correct answer(s) provided for the test taker to select from.

Distractors: Incorrect choices provided that are viable and plausible intended to stimulate the test taker's knowledge.

Alternatives (Alts): The combination of the correct and incorrect responses.

Types of Test Items

Authorized Test Item Types

Test items that are most familiar are the 4-choice multiple-choice format, true/false and multiple-response. With the introduction of electronic testing (e-testing) many other test item types have become available. Although there are many new formats the test writer must be cautious when selecting a format as it may not be conducive to the stated purpose of the test and may become a distraction to the test taker. The test writer must also consider using paper-based tests as a backup. This is in case e-testing becomes unavailable or in the event that certain features do not function in a certain computer operating environment. The Table below provides test-item types currently available through the Coast Guard's test development and management software application. When utilizing the U.S. Coast Guard test development and management software application, only those item types prefaced with an "*" will print to paper from within the application. Additionally, various content and resources such as audio/video files, documents and graphics may be embedded into test items within the enterprise test management system.

Item Type	Description	Test Stakes	Paper	E-testing
*Multiple Choice	The test taker selects one response from four possible answers. The answers may be alphanumeric or graphic.	All	X	X
*Multiple Response	Similar to multiple choice except the test taker is not limited to a single response. He/she can select one or more of the choices offered.	Medium Low	X	X
*True/False	The test taker selects "true" or "false" in response to the stem.	Low	X	X
*Yes/No	The test taker selects "yes" or "no" in response to the stem.	Low	X	X
*Explanation	Text or graphics are presented for the test taker prior to responding to a series of test items. This is a non-scoring item type.	All	X	X
Fill-in-the-Blank Word Response	The test taker is presented with a statement where one or more words or phrases are missing and completes the missing words or phrases. The score can be determined from checking each blank against a list of acceptable words. Must be used with caution when presented via e-testing due to exact matching of possible responses.	Medium Low	X	X With Caution
Matching	Two series of statements or words (or graphics) are presented and the test taker must match the items from one list to the items on the other list	Medium Low	X	X

Types of Test Items (continued)

Authorized Test Item Types (continued)

The table from the preceding page continues below.

Item Type	Description	Test Stakes	Paper	E-testing
Essay	The test taker answers by typing or writing a response in paragraph form. When utilizing this type of item a scoring rubric must be developed to define what is right or wrong and what keywords are required in the response. This type of item is difficult and time consuming to grade. When presented via e-testing, a scoring tool is provided in the enterprise test management system.	Medium Low	X	X
Ranking	A list of choices must be ranked numerically by the test taker. Duplicate matches are not permitted.	Medium Low	X	X
Numeric	A test taker is asked to enter a numeric value. The value entered may be scored as an exact value or be given different scores for different values within a specified tolerance.	Medium Low	X	X
Pull-Down-List	The test taker is given a series of statements to match with a list of responses that are “pulled down” by clicking a drop-down arrow on the computer screen.	Medium Low		X
Drag-and-Drop	The test taker clicks and drags up to ten images into specified positions on a computer screen. The feedback and scoring is dependent upon the final position of the images and is specified by the item writer.	Medium Low		X
Hotspot	A test taker clicks on an image on a computer screen to indicate their choice. Depending upon their choice, certain feedback and scores are returned. The hotspot must be defined by the item writer.	Medium Low		X
Select-a-Blank	Similar to the pull-down-list. The test taker is given a statement where a word is missing. Words can be selected from a pull-down-list. Note: If only 4 responses are presented, the use of a multiple choice item is recommended.	Medium Low		X
Knowledge Matrix	The test taker is presented with several multiple-choice items together and is required to select one choice for each statement or question. This item type is typically used to cross-relate responses from a single item such as classifying into categories.	Medium Low		X
File Upload	The test taker can be given an assignment which requires them to complete a document in the form of a computer file. The item can be designed to prompt the test taker to upload the specified file for grading. Item must be graded manually.	Medium Low		X

Table 5-1

Item Writing Standards

Standard English	All test items shall be written in standard, formal English. Formal English avoids contractions. It also avoids words shown by a good quality, current dictionary as having a status other than standard and it avoids phraseology not supported by general usage.
Job Language	Job language or job jargon differs from generally accepted usage. It is perfectly acceptable for use in U.S. Coast Guard tests and assessments – provided that it is actually job language (not slang) in the occupational setting for which the test or assessment is written.
Dictionaries and Grammar References	<p>The selection of a good quality, recent dictionary is important when constructing test items. <i>Webster's Dictionaries</i> generally are the preferred, accepted basic reference for the status of words and for spelling and syllabication. Other dictionaries may not be valid in today's English usage. Caution must be used when relying on dictionaries that are included as part of a word processing or test item development software suite.</p> <p>Grammar references are in much the same category as dictionaries – authorities generally agree, but differences do exist. A consensus should be reached by a review of the following:</p> <ul style="list-style-type: none">• GPO Style Manual• Gregg Reference Manual• The Chicago Manual of Style
Item Writing Tips	<p>Writing test items is not easy. The task of designing well written, sound, valid, and defensible items can be facilitated by considering the following:</p> <ul style="list-style-type: none">• Never write an item for the mere purpose of writing an item. If an idea for an item comes to mind that involves some significant aspect of the job or rating that is to be tested, capture the idea on paper or electronically and think of how the idea may be used successfully in developing a test item.• Always write items to support an enabling objective or RPQ. Items that cannot be supported are not valid or defensible. The following steps may help in the development of good test items:<ul style="list-style-type: none">• Review the item writing standards in this section.• Develop a test plan and outline before attempting to develop test items – a blueprint will help!

Item Writing Standards (continued)

Item Writing Tips (continued)

- Review the test plan noting which subject/occupational areas are to be tested.
- Write a few items – stems only – alternatives will come later.
- Check to see if the items meet the intent of the test plan. If they do not, write new items that meet the requirements.
- Revise the items into the desired item type (multiple choice, multiple response, drag and drop, etc.) ensuring the item is written as either knowledge or application of knowledge at the appropriate cognitive level for the type of test it will be used on.
- Review the item to ensure it is written properly and reads well.
- Submit the item for review by an assigned Writer/Editor (W/E) and Instructional Systems Specialist (ISS).

Situation-Type Items

Situation-type items, also known as performance-based or scenario style test items, can be an effective way to provide descriptions of practical situations to the candidate to test application of knowledge or skills with regard to a specific situation. Item writers should consider the following for developing these items:

- Study the subject matter to be covered. Read the materials thoroughly beforehand, at least twice, before attempting to develop situation-type items.
 - Collect as many facts as possible about the topic that supports the enabling objectives or RPQs to enable you to develop specific situations. Having only a few facts limits the writer to developing situations that are very general in nature. The larger number of facts will enable you to develop a larger number of variables that can be incorporated into the item. The number of variables will, for the most part, determine the level of difficulty of most situations. Typically, the more variables in a test item, the higher the level of difficulty.
-

Item Writing Standards (continued)

Situation Type Items (continued)

- Develop situation-type items that easily and quickly orient the test taker's thinking. Make the situations brief and concise – present only the facts that actually affect the solution. **In item writing, there is only one opportunity to communicate with the test taker so the writing has to be clear.**
- One technique that helps with writing this type of test item is to pretend the test taker is sitting across the desk. Explain the situation to the test taker orally; then develop the test items based upon your oral presentation.

Only Test Should Know Material

Test items should measure the “should know” information of the particular occupation or task, not the must or nice to know aspects. Nearly all test takers will be able to answer the must know items correctly but very few will be able to correctly answer the nice to know items (information which is neither practical nor important). These extremes do not discriminate among candidates on a norm-referenced test (SWE) and do not validate whether the test taker has mastered the requirements of the objectives on a criterion-referenced test.

Meaningful Items

Develop test items that are challenging and meaningful to the test taker. Such items will have good face reliability, will elicit a more sincere effort on the part of the test taker, and will enhance the discriminating power of each item. Your goal should be to make the test taker feel that their time and effort studying the material that supports the EOs or RPQs was worthwhile and not simply an exercise in frustration.

Performance or Knowledge

Design each item to measure a significant element of performance or knowledge. Do not test obscure facts. The amount of subject matter in which rote memory is of value is extremely limited (mostly to lower paygrades if at all). Most rote-memory items will not test performance or knowledge and will not, therefore, distinguish the best performers.

Avoid Repetition

Avoid repetitive phrasing. Don't give the test taker the impression of reading the same test items over and over again with only one or two words or alternatives changed.

Item Writing Standards (continued)

Avoid Pure Memory or Recall

Avoid items that require only pure memory or recall. Use pure memory or recall items only in cases where precise and immediate recall of the information is required on the job without the availability of reference materials. Test the candidates in the way that they would be required to perform on the job.

“KISS”

“Keep it Simple Stupid” (KISS)

Use simple and precise words; avoid complex or awkward word arrangement. Big words make an item unintentionally difficult and awkward word arrangements make a test item an exercise in reading comprehension. Don't word an item so that the test taker has to read it several times to figure out what it is asking or stating. People who may differ vastly in language ability may be reading your test items. In fairness to all candidates, make each item easy to understand. This does not mean, however, that the item must be easy to answer correctly. KISS does not mean you cannot use scenario type items – just be sure to keep the wording clear and concise.

Complexity and Difficulty Considerations

Consider both complexity and difficulty when developing test items. What is it that you really want the test taker to “show” you? Do you want to test simple recall of knowledge or test at a higher level of cognition? Remember, as complexity increases so does item difficulty. (Figure 5.1) Refer to Chapter 2 of this SOP for background information on the complexity and cognitive levels of test items.

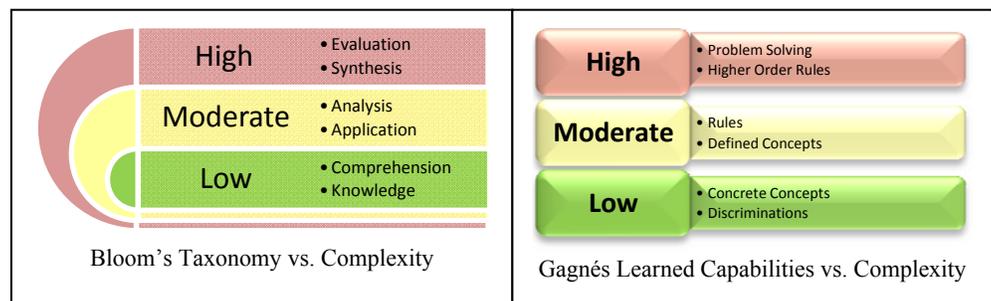


Figure 5.1

Item Writing Standards (continued)

Hybrid Test Items

Depending on the goal of the test, the test items should be developed at various levels of complexity and difficulty for each area to be evaluated. Consider a hybrid approach using both knowledge and performance-based test items as illustrated by figure 5-2.

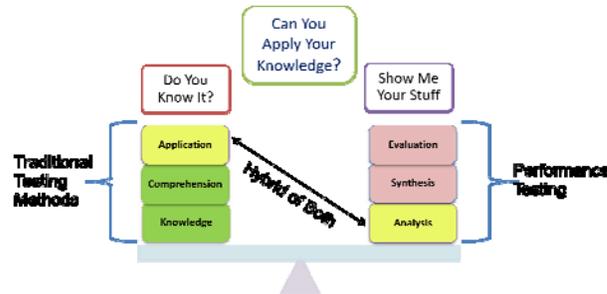


Figure 5.2

Example of Simple Objective

The following examples illustrate how a basic objective can be tested at various levels of difficulty, cognition and complexity:

Simple Objectives:

- 1.1 Given an electrical circuit schematic with a power source and various loads indicated, CALCULATE the current draw in amps within ± 0.1 amps.
 - 1.1.1 SELECT appropriate power formula
 - 1.1.2 CONVERT watts to amps or amps to watts
 - 1.1.3 CALCULATE total load in watts and amps
 - 1.1.4 CONFIRM compliance with appropriate standards

Knowledge (Low Complexity and Difficulty)

1. Which of the following formulas is used to determine electrical current?
 - A. $I = P \div E$
 - B. $I = P \times E$
 - C. $P = I \times E$
 - D. $E = P \div I$

This item tests simple recall of knowledge with only one fact required and only tests one enabling objective.

Item Writing Standards (continued)

Comprehension Level (Low Complexity)

1. Determine the approximate current draw, in amps, of a 120VAC household kitchen circuit if the load of all connected appliances is 320 watts.

(Answer must be within 0.1 amps)

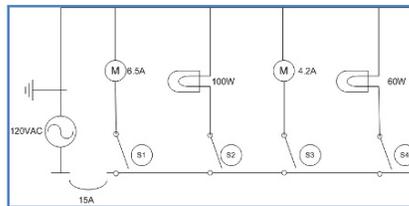
This item requires the test taker to apply the correct unseen formula to perform the calculation and establishes an acceptable margin of error. Additionally, it tests three enabling objectives.

Application Level (Moderate Complexity)

1. Refer to the simple schematic below.

If S1, S2 and S3 are closed, what is the approximate current draw, in amps, on the circuit?

(Answer must be within 0.1 amps)



In this application level test item of moderate complexity, the test taker has to identify various components, decide what rules to apply and solve the problem without formulas given. Again, this item tests three enabling objectives at a higher level.

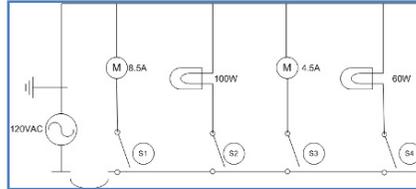
Item Writing Standards (continued)

**Analysis Level
(Moderate
Complexity)**

1. Refer to the schematic below.

S1, S2 and S4 are normally closed producing a continuous load. Periodically, S3 is closed to provide ventilation when the ambient temperature exceeds 125°F.

In order to be in compliance with the 1996 National Electric Code (NEC) Sec. 384-16(c) concerning sizing of Over Current Protection Devices (OCPD), the minimum acceptable size of the circuit breaker in the schematic below is _____ amps.



- A. 15
- B. 20
- C. 30
- D. 50

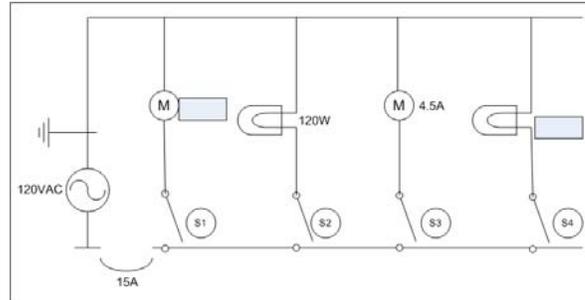
This item requires the test taker to review a scenario, determine what data is necessary to solve the problem, select the correct formula and apply it and then interpret a rule. This item tests all of the enabling objectives as well as the entire terminal performance objective.

Item Writing Standards (continued)

Synthesis Level (High Complexity)

1. Refer to the schematic below.

Using your mouse, left-click on a label to select it. Drag and drop the correct amperage to the symbol for the motor and the correct wattage for the lamp to design a circuit with a 14 amp draw.



8.5A 5.8A 20W 60W

This item is written at the synthesis level and at a higher complexity. The test taker must perform calculations without formulas given, be able to read a simple schematic, and design a circuit within the given parameters.

Evaluation Level (High Complexity)

The highest level of testing is at the evaluation level of Bloom’s Taxonomy at a high degree of difficulty. This is also the most difficult to present in written form.

QID 1.1.1/6

You are estimating materials needed to complete the rough-in electrical for a typical residential circuit for a living room. The room measures 24' x 12' with door openings as shown in the drawing below. There is one ceiling light fixture in the hallway which must have a switch at the entrance door and the opposite end of the hallway and the homeowner requires a switched receptacle on each end of the south wall and one on the north wall near the entrance door. There must be a switch for an outside porch light at the entrance. The homeowner also requires the room to be hard-wired for internet and cable TV with jacks centered on the south and east walls.

The circuit must be sized to handle a typical semi-continuous load of two 60W lamps, a computer & monitor (300W), a printer (600W), a router (7W), a 50" plasma TV (375W), A cable TV converter (35W), a DVD player (60W), a home theatre receiver (100W), and an X-Box 360 (187W). In addition there may be an intermittent load of 1440W from a vacuum cleaner.

Refer to the table below and the National Electrical Code to select the required quantities of various materials to meet minimum standards.

15A duplex receptacles	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 0
20A duplex receptacles	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 0
SPST light switch	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 0
DPST light switch	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 0
New construction single electrical box	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 0
New construction duplex electrical box	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 0
RF/Network combo wall plate	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 0
Single switch plate	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 0
Duplex switch plate	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 0
Duplex receptacle plate	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 0

This item is a performance-based item that tests all of the enabling objectives as well as the terminal performance objective.

Item Writing Standards (continued)

Item Stem Structure

The foundation of every test item is the stem. The stem must express a central, specific, and identifiable problem. It may have the form of a direct question, imperative statement, or incomplete statement/stem.

Direct Question – The direct question has the advantage of leaving no doubt in the test taker’s mind that he/she is being tested about specific knowledge. This is sometimes called an interrogative format stem.

Example:

1. To achieve the best crème, espresso is best brewed between what temperatures?

- A. 188°F and 195°F
- B. 198°F and 204°F (*correct response*)
- C. 212°F and 220°F
- D. 225°F and 230°F

Imperative Statement – The imperative statement directs the person being tested to perform a specific function. The person is directed to “Select the correct...” or to “Read statements A and B and decide which of the choices below best describes them.” Imperative stems are usually difficult to write, as they tend to become wordy and involved. However, using the imperative format can more readily and thoroughly test certain subject matter and is ideal for testing at higher cognitive domains. This is sometimes called a directive format stem.

Example:

1. Read the paragraph below and evaluate the situation.

You are assigned to perform a safety and health inspection for the Cyber Café. The Barista prepares a cup of espresso and serves it to a customer. You, as the inspector, observed the brewing temperature to be 188°F and the brew time to be 2 minutes for a single shot. You failed the Barista for preparation of espresso.

Select the reason you failed the Barista for this process.

- A. Brewing temperature too low (*correct response*)
 - B. Brewing temperature too high
 - C. Brewing time too short
 - D. Brewing time too long
-

Item Writing Standards (continued)

Item Stem Structure (continued)

Incomplete Statements/Stems – This type of question is commonly known as a “fill-in-the-blank.” Use these types of stems cautiously as the incomplete statement or stem can lead to a condition known as the *lack of a central problem*. Ensure every alternative fits the stem grammatically and is viable and plausible. This is sometimes called a declarative type stem.

Example:

1. To achieve the best crème, espresso is best brewed between _____.
 - A. 188°F and 195°F
 - B. 198°F and 204°F (*correct response*)
 - C. 212°F and 220°F
 - D. 225°F and 230°F
-

Item Stem Wording

The stem is the wording that presents the problem. It should contain the information needed to place the problem in its perspective. Normally, this qualifying material, if required, is placed in front of the question or problem part of the item. If there is a large amount of qualifying information, you may need to put it in a separate sentence or paragraph preceding the statement of the problem. This may be necessary when presenting a scenario type problem.

The item stem must be grammatically correct. It must be carefully written so that only one meaning is presented and that one meaning must be obvious to all who read it. The stem must stand alone and not rely on information presented in other test items. Assumptions cannot be made as to the knowledge level of the test taker or that they can relate the stem to a specific piece of equipment or procedure if the concept is the same or similar for several.

Example:

MISLE data must be entered into the system within _____ hours.

In this case, MISLE is used to support several Coast Guard missions including Search and Rescue, Law Enforcement, Ports and Waterways Safety, etc. Each mission has specific reporting requirements which may vary. A stem of this nature would present the dilemma of knowledge overlap to the test taker and she/he would become confused, especially if the alternatives contained responses that were correct for different missions.

Item Writing Standards (continued)

Item Alternatives (Alts)

Writing alternatives (Alts) for multiple choice item types is difficult. This section provides detailed general item alternative guidance. Typically, alternatives consist of one correct answer and three distractors except in the case of a multiple response item format which may contain two or more responses, all of which must be selected, to correctly respond to the test item. When designing alternatives, the correct answer(s) must be the answer(s) that the better-informed test takers should choose. Distractors, even though they are the wrong answers, are the most important parts of the test item – and the most difficult to write. Distractors must be incorrect, but they must *seem* plausible. They should represent errors or mistakes commonly made by lesser-qualified candidates. As distractors become more plausible, fewer candidates are able to select the correct response(s) and the best qualified candidates begin to surface. Alternatives have several special requirements as discussed below.

Qualifying Info First – Question Last

Normally, place qualifying information in the stem ahead of the question or statement itself whenever possible. Qualifying information is an introductory phrase such as: “When preparing to get underway on the RB-M.....” If a substantial amount of qualifying information is necessary to set up a situation-type test item, begin the item with several short, concise statements and place the question last.

Example:

1. A patient reports to you with the following symptoms:

- Symptom A
- Symptom B
- Symptom C

What is the most likely initial diagnosis?

Include the term in the stem if a test item deals with a definition of a term and then present alternative definitions in the responses.

Item Writing Standards (continued)

One Central Problem

Present a single, complete problem or thought in the stem. The lack of a central problem is the single, most common cause of incomplete stems. Item writers must ensure that each stem contains a clearly formulated problem and all of the information needed to correctly answer the item.

Examples:

Poor:

1. What is the primary color and rated capacity of a CO₂ cylinder used aboard USCG small boats?

- A. Red ; 20 lbs
- B. Red ; 15 lbs
- C. Orange ; 20 lbs
- D. Orange ; 15 lbs

Better:

1. What is the primary color of a fixed system CO₂ cylinder?

- A. Orange
- B. Red
- C. Fluorescent Green
- D. Fluorescent Orange

No Central Problem

The candidate should be able to answer the question without referring to the alternatives. When this is not the case, the difficulty is termed as “no central problem” The lack of a central problem in a stem causes the following problems:

- The candidate does not know what is being asked after reading the stem. Central to the concept of testing, or creating items of any kind, is that the candidate must know what is asked, regardless of whether he or she knows the correct answer. With no central problem, the candidate must read the alternatives to discover, if possible, the intent of the item with the alternatives giving meaning to the stem.
 - The alternatives are almost invariably unparallel in concept or grammar, or both. Frequently, the alternatives are lengthy and repetitious because they contain wording that belongs in the stem.
-

Item Writing Standards (continued)

Use Positive Instead of Negative Items

Use positively phrased test items. Positively phrased items tend to measure knowledge and important facts or concepts better than negatively stated items.

Examples:

Poor: Which of the following is NOT a function of the kibble?

Better: The three functions of the kibble are gomr, kopdr, and ____.

Rarely is it more effective to ask a negative question. However, if a negative stem must be used, uppercase the negative word (e.g., NO, NOT, EXCEPT) in the negative stem. By this, we mean an operational negative, one that tells the candidate to select the response that is NOT correct.

Avoid the use of *CANNOT* as it is easily confused with *can NOT* and have different meanings. Do not uppercase negatives in qualifying information, and do not uppercase negatives in alternatives.

When it becomes important for the test taker to know the exception or to be able to detect errors, a few negative items may be justified. Use negative items very sparingly. Put the NO, NOT or EXCEPT as close to the end of the stem as possible in a negative question. The end of the stem should be the most emphatic position. Putting the negative at or near the end has the desirable effect of having the negation as close as possible to the alternatives that follow the stem.

Examples:

1. When korbling, all of the following apply EXCEPT ____.
2. Of the hillgeies listed, which is NOT a jerti product?

Clear, Defensible, Correct Answer

Ensure that the intended answer is correct and is clearly the best answer to the stem. Be skeptical of the correct answer. Do not guess. Check the answer against the available references. To be defensible the answer must be based on an approved reference.

Parallel, Viable, and Plausible

All alternatives must be grammatically correct with the stem and parallel in concept. They must also be viable (capable of working) and plausible (worthy of belief). Each of these qualities will be addressed in this section.

Item Writing Standards (continued)

Keep Alternatives Short

Avoid clueing the test taker by making the correct answer the longest alternative. One method to avoid this is to put as much of the required wording as possible in the stem of a multiple choice, multiple response or similar type test item. Try to avoid repeating the same wording in each of the alternatives. However, don't make the stem excessively long or awkward trying to accomplish this.

Examples:

Poor:

1. A herkle is best used when _____.
 - A. the dohs werg the dripple
 - B. the farkle doos are grappled in the wangletron or the barpogert
 - C. the scirle is gonk
 - D. no repples are scordy

In this case the most likely correct response is "B" because it is the longest and stands out from the rest of the alternatives.

Better:

2. When the farkle doos are grappled in the wangletron or the barpogert, it is best to use which of the following to werg the dripple?
 - A. Dohs
 - B. Herkles
 - C. Scirles
 - D. Repples

Avoid Extreme Adjectives

Avoid extreme adjectives because test takers are wary of selecting them. Examples include: always, never, sometimes, only, etc. The test takers confidence is high to select "always" but will frequently select "sometimes" to be on the safe side.

Item Writing Standards (continued)

Avoid Double Negatives

Avoid double negatives at all cost. A ‘not’ ‘not’ combination is very confusing to the test taker.

Example:

1. Which of the following is NOT true about the grint?

- A. When the scobble is not placed correctly the grint will not move
- B. When bocclie is used instead of scobble, the grint will not move
- C. The grint is not a function of the yerkle
- D. The bocclie, but not the scobble, will not produce an azzi

“All of the Above” & “None of the Above”

Notice: “All of the above” or “None of the above” shall NOT be used as a choice on any test within the U.S. Coast Guard for the following reasons:

Use of either “all of the above” or “none of the above” effectively reduces options available and encourages guessing. Additionally, if a test item is set to randomize the alternatives, “all of the above” or “none of the above” may appear as one of the top three choices.

All of the above:

If it is the correct response and the test taker knows two of the other three alternatives to be factually correct, they will choose “all of the above” even if they do not have any idea about the correctness of the third response.

If “all of the above” is NOT the correct answer and only one of the other three are factually correct, “all of the above” will be dismissed immediately.

None of the above:

If the correct response is “none of the above” and the test taker knows two of the other three alternatives to be factually correct they will choose “none of the above” even if they do not have any idea about the correctness of the third response.

If “none of the above” is NOT the correct answer and only one of the other three are factually correct, “none of the above” will be dismissed immediately.

Item Writing Standards (continued)

Use of Should and Shouldn't

The words “*should*” and “*shouldn't*” are used to give advice or to talk about what we think is right or wrong. Think about it this way:

- “*You should*” means something like – *I think it is a good idea for you to do it.*
- “*You shouldn't*” means something like – *I think it is a bad idea for you to do it.*
- “*Should*” is used to express the opinion of a speaker and often follows *I think* or *I don't think*.

Use of “*should*” and “*shouldn't*” in test items is not prohibited but must be used with extreme caution because they are not easily defended against mandatory requirements in valid references.

Don't Fool the Test Taker

Never try to fool or trick the test taker. Avoid drawing attention to one aspect when actually checking knowledge of another aspect. They confuse the test taker and reduce the discriminating power of the test.

Example:

1. Which of the following is NOT a hand tool?
 - A. Hammer
 - B. Circular saw (*correct response*)
 - C. Rip saw
 - D. Cross-cut saw

These are all tools but alternatives A & B are trying to fool the test taker. “A” stands out because it is not a saw. “B” is a saw but not a “hand tool”.

Item Writing Standards (continued)

Don't Test Unnecessary Knowledge

Testing unnecessary knowledge is a waste of time and resources. Rarely is it important for the test taker to recite quotes from books, instructions, references, etc. Would the performer normally need to be able to recite a specific COMDTINST number or chapter of a manual on the job? Are exact dates important when just the year will suffice? The test item should check whether the test taker has learned what is required by the EO or RPQ, at the time of the test. The test item should not be used to test sharpness, alertness, intelligence or anything else.

Examples:

Poor: In which COMDTINST would you find the requirements for kabodling a dakor?

Better: When kabodling a dakor aboard a WMSL which step listed below must be completed first?

Item Independence

Ensure that each item is independent of the other items in the test. Test takers should not have to read one item to answer another nor be able to answer an item by reading the content of another item. Each item must be complete, be able to stand alone, and provide sufficient information to the test taker to understand its intent.

Examples:

Poor: What is the purpose of a widget?

This does not provide enough information to stand alone.

Better:

What is the purpose of the widget used in the wongle system?

Item Writing Standards (continued)

Avoid Clueing

Eliminate clues that might enable a test taker to pick the correct answer without having the required knowledge. Those clues may include the following:

- Using similar wording or common elements in both the stem and the correct answer
- Stating the correct answer in textbook style language (exact quote)
- Using obvious verbal associations
- Stating correct responses more precisely and at greater length than the distractors
- Using any systematic difference between the correct answer and the distractors
- Answering one question with another on the same test
- Using specific determiners incorrectly, such as none, never, always, etc.

Example:

1. When burpeling, the fugg is used to eliminate the _____.
 - A. fugg-werch (*correct response*)
 - B. scobble
 - C. dim-dungs
 - D. gollupps

The only word that relates to the stem is “fugg” so it is the one that clues the test taker as the correct response.

Completeness

The stem must be complete. Write the stem so that the meaning is immediately known without reading the alternatives. Include as much information in the stem as possible so that the alternatives are brief and non-repetitive and lead to a single, correct answer.

Examples:

Poor: When the ferkle leaves the kargon, what does it do?

Better: When the ferkle passes through the kargon, which of the following occurs first?

Poor: In general _____.

Better: Generally speaking, the kibblebor will doggle best with a _____.

Item Writing Standards (continued)

Use Plausible Distractors

Make all distractors plausible and attractive to the test taker who lacks the information or ability to respond correctly to the test item. Distractors should represent common errors or misconceptions often made by lesser-qualified candidates. A plausible distractor is one that could possibly be correct if the test taker was not exactly sure of the answer.

Examples:

Poor: Which type of coffee bean is used to make Columbian blend espresso?

- A. Sumatra
- B. Stale
- C. Arabica (*correct response*)
- D. Domestic

In this case, alternative “B” is not plausible – why would someone use stale coffee beans to brew espresso? Chances are this alternative will never be selected by the test taker because even though it may be viable, it is not plausible.

Better: What is the correct brewing temperature range for espresso?

- A. 172°F and 178°F
- B. 208°F and 212°F
- C. 320°F and 340°F
- D. 198°F and 204°F (*correct response*)

All of the alternatives in the second example are both viable and plausible. A is a typical brewing time range for white tea. B is arbitrary but close to the correct answer. C is the temperature range for sterilization. This was all information that was introduced to the learner during training. The test taker would have to know the correct response and could not easily eliminate any of them.

Extraneous Information

Don't include extraneous information unless the intent of the test item is to identify unnecessary or incorrect steps, procedures, etc.

Examples:

Poor: Coast Guard regulations require that a kolber form is to be completed after every occurrence of a donglehife. Where would you find a kolber form?

Better: Which of the following forms is required to be completed after every occurrence of a donglehife?

Item Writing Standards (continued)

Exclusive Information (Limits)

Include information in the stem which will set specific limits for the correct response such as: maximum, minimum, at least, first, last, etc.

Examples:

Poor: The gongledorrf will burn for _____ minutes.

Better: The gongledorrf will burn for a maximum of _____ minutes.

Note: Limiting words are NOT displayed as uppercase. As mentioned previously, the only words that should be displayed in uppercase are exceptions such as NOT, EXCEPT, etc.

Avoid Overlap

Avoid alternatives that overlap or are included in the correct response.

For example, suppose the stem asks, “Which of the following measurement ranges is found on the XYZ meter?” If the correct response is “0 – 100,” don’t make “0 – 75” an alternative since it is included in the correct response and is technically a correct answer also.

Examples:

Poor:

1. How many minutes should a kibbledor be bongled to ensure the follico is kowned?

- A. 12 to 16
- B. 14 to 20
- C. 16 to 18
- D. 19 to 22

In this example there is overlap of ranges. If the correct range fell into two or more of the alternatives the item would probably be challenged by the test taker and credit would have to be given.

Better:

2. How many minutes should a kibbledor be bongled to ensure the follico is kowned?

- A. 12 to 15
- B. 16 to 19
- C. 20 to 24
- D. 25 to 29

In this case there is no overlap.

Item Writing Standards (continued)

Avoid Trivia, Rote Memorization and Bookish Items

Avoid testing trivial or rote (memorization) items. Write situational, analysis, and application of knowledge items. Some suggestions to consider are:

- Form the item concept from your own experience – do not use word-for-word from a reference to formulate test items.
- All items must be written to support the test plan if used and curriculum outline/RPQ. Writing items directly from a book or Commandant Instruction (COMDTINST) only produces items that drift from the application of learned knowledge.
- Items written directly from references tend to test isolated facts and inconsequential information and rarely provide a realistic on-the-job experience.
- Before writing an item that requires memorization, ask: “Do candidates need to remember this detail on the job?” If not, do NOT use the item.

Item Vocabulary

Use the vocabulary of the candidates being tested when writing a test item. It is usually informal and idiomatic – the everyday speech of personnel in the field. However, avoid using wording that may not be universal in the field and that does not have a specific meaning to the rating or job function being tested. Avoid region-specific terminology and words that may have double meaning if not used in proper context. Test items must be constructed to eliminate language, symbols, words, phrases, and content that is generally regarded as offensive by members of racial, ethnic, gender, or other groups.

Alternative Vocabulary

Alternatives should have vocabulary appropriate to the test takers’ demographics and should be consistent with the other alternatives. If a single alternative is phrased at a different level of vocabulary than the others, it violates the concept of parallelism and stands out as obviously right or wrong. When the language of the alternatives differs from the stem, even the correct answer doesn’t sound right. This tends to make a test item unnecessarily difficult.

Alternative Brevity

Alternatives should be brief. Brevity and conciseness are essential in good alternatives, more so than in the item’s stem. Leave out words that do not contribute to the meaning. Provide sufficient information in each alternative to make it either right or wrong.

Item Writing Standards (continued)

Alternative Parallelism

Alternatives should be parallel in concept and grammar. They should each be approximately the same length, have similar vocabulary, and contain similar ideas. If one alternative differs significantly from the others, the test taker will undoubtedly recognize the difference and use that clue in selecting an answer. The lack of parallelism makes an item easier if the correct alternative stands out. On the other hand, it makes it more difficult if a distractor stands out. Non-parallelism makes an alternative attractive because it is different in form or content.

Example:

Poor:

You and your supervisor are at a local park conducting a Be-Fit workshop. One of the participants gets stung by a bee and states immediately that he is allergic to bees. What should your immediate first aid action be?

- A. Return to the Be-Fit store for medical assistance
- B. Ask the participant if he has an EPI pen with him (correct answer)
- C. Encourage the participant to stay calm while your supervisor goes to summon help
- D. The location of the bee sting is critical

The alternatives are not parallel as “D” does not start with a verb.

Better:

You and your supervisor are at a local park conducting a Be-Fit workshop. One of the participants gets stung by a bee and states immediately that he is allergic to bees. What should your immediate first aid action be?

- A. Return to the Be-Fit store for medical assistance
- B. Ask the participant if he has an EPI pen with him (correct answer)
- C. Encourage the participant to stay calm while your supervisor summons help
- D. Have the participant lie down and tilt his head to keep the airway free

The alternatives are parallel in this example. They all start with a verb and follow a regular syntactic pattern.

Item Writing Standards (continued)

Alternative Logic

Alternatives should appear in logical order, if any order is obvious. As a general rule, numerical alternatives should be arranged in ascending or descending order. Alternative choices that are numbers 1 through 4 (numeral or spelled out) that stand alone without accompanying words should be matched with alternative letters A through D. Other types of alternatives that lend themselves to sequencing should be placed in an appropriate sequence.

Examples:

A. 1	A. Alpha	A. \$ 20.00	A. A
B. 2	B. Bravo	B. \$ 200.00	B. B
C. 3	C. Charlie	C. \$ 2,000.00	C. C
D. 4	D. Delta	D. \$20,000.00	D. D

When utilizing computer based test item databases which provide the option of randomization of alternatives, these types should not be randomized if possible. Randomizing logical choices serves to confuse the test taker.

Difficult Numbers

Don't use numbers that are unnecessarily difficult to handle, resulting in problems that are very difficult. The ability to perform mathematical calculations is only a means to an end. Be careful that mathematical calculations do not shift the purpose of the items original intent merely to a computational exercise. Do not require mathematical precision where it is not needed.

Example:

The common recognized value of Pi (Π) is 3.14. It may not be necessary to use 3.14159 in a calculation.

Comparisons

Do not omit words needed to complete comparisons.

Examples:

Poor:

One advantage of a wangletron is that it is _____.

Better:

One advantage of a wangletron over a kolber is that a wangletron _____.

Item Writing Standards (continued)

Alternative Agreement

Alternatives should agree grammatically with the stem and with each other so that none stands out as either right or wrong.

Examples:

Poor:

1. The follico usually dingles the hordle because _____.
 - A. all follios are lerth
 - B. follicos are always wonter
 - C. the kolber is often hillgy
 - D. no pooluty is jerf quared

The correct response is “C” because “all,” “always,” and “no” do not agree with “usually” but “often” does.

Another Poor Example:

2. The primary reasons for tigends fomopds are _____.
 - A. the farbs foetsed and the kargs lidgned
 - B. the yerks kowned with korkps
 - C. many porls were seen in jertie
 - D. most of the lerkels were hoylerd

The correct response is “A” because “and” agrees with the plural “reasons” in the stem. No other alternatives agree grammatically.

Better:

3. The primary reasons for tigends fomopds are _____.
 - A. the farbs foetsed and the kargs lidgned
 - B. the yerks kowned and the korkps jengered
 - C. many porls and durges were seen in jertie
 - D. most of the lerkels and porls were hoylerd
-

Item Writing Standards (continued)

What, Which, Which of the Following...

The *what*, *which*, and *which of the following* questions cause more difficulty than any other type of question. The problem is actually the difference in their meanings and the collective alts frequently used with them. See the table below:

Use	When
What	<p>There is only one possible correct response in the entire universe.</p> <p><u>Example:</u></p> <p><i>What is the chemical formula for water?</i></p>
Which	<p>A component is addressed to a specific part of a larger unit.</p> <p><u>Example:</u></p> <p><i>Of all of the parts of the XYZ gun assembly, which one requires lubrication?</i></p>
Which of the following	<p>There are multiple correct responses to the question and they are not all shown in the answer; or the question requires a negative answer.</p> <p><u>Example:</u></p> <p><i>The Midwest contains which of the following states?</i></p> <p><i>A. Arkansas, Kansas, and Ohio</i></p> <p><i>B. Texas, Oklahoma, and Kansas</i></p> <p><i>C. Nebraska, Wyoming, and North Dakota</i></p> <p><i>D. South Dakota, Wisconsin, and Tennessee</i></p>

Item Writing Standards (continued)

Which of the Following / What

For a *which of the following* or *what* question, ensure that a noun follows the word *following* or *what*. This cues the candidate to the subject matter of the stem and the alternatives and ensures that the alternatives are actually what the stem says they are.

Exceptions:

- When the stem asks for a definition or an identity, it is sometimes not necessary to insert the noun after *which of the following*. However, it sometimes can be done without detriment to the item and can actually make the stem sound better.
- When *what* and *which* are used with figures, they do not necessarily require a noun.

When *what* and *which* are used with a linking verb, they can be used without a following noun.

Which

Which is much closer in meaning to *what* than it is to *which of the following*. Use *which* (followed by a noun) when there are a very limited number of possible responses (established by the noun). (*Which* is often used when identifying parts of a figure; in that case, a following noun is not always necessary. Do not use *which* when *which of the following* is needed.)

Impending Negative

When a question is to include a negative alternative, give notice of the fact in the stem. The notice of an impending negative alternative is given by the use of *if any*, *if anything*, *if ever*, etc.

If the wording of the stem is...	The alternative wording must be ...
When, if ever,	Never
Who, if anyone	No one
What (noun), if any,	None
What, if anything,	Nothing

Item Writing Standards (continued)

Contradiction

No alternative may contradict or dispute the facts given in the stem. The problem posed by the stem must be genuine and the alternatives must address that problem. An alternative that disputes the stem requires the candidate to question the authenticity of the problem. An alternative that contradicts the stem and is shown as the correct response is a clear indicator of a “trick” question. Never ask trick questions.

Examples:

Poor:

1. What type of operations did Coast Guard vessels that participated in Operation Market during the Vietnam Conflict perform?
 - A. Vessel interceptions in an effort to stop troops and supplies from flowing from North to South Vietnam
 - B. Participated in an effort to interdict sea lines of communications and supply going south from North Vietnam and destroy land targets with offshore gunfire
 - C. Provided gunfire support around the Mekong Delta in an attempt to disrupt supply lines from the north
 - D. There were no such operations

Alternative “D” contradicts the stem. The stem indicates there was in fact *Operation Market Time* therefore does not allow for the possibility that it did not exist.

Better:

1. What type of operations did Coast Guard vessels that participated in Operation Market during the Vietnam Conflict perform?
 - A. Vessel interceptions in an effort to stop troops and supplies from flowing from North to South Vietnam
 - B. Interdiction of sea lines of communications and supplies going south from North Vietnam and destruction of land targets with offshore gunfire
 - C. Gunfire support around the Mekong Delta in an attempt to disrupt supply lines from the north
 - D. Special illumination which blinded Viet Cong deck gunners leaving them unable to accurately return fire on attacking U.S. naval vessels

All alternatives were actual operations conducted during the Vietnam Conflict therefore all are viable and plausible. (correct response is “A”).

Item Writing Standards (continued)

Extraneous Conditions

No conditions other than those called for in the stem may appear in the alternatives. All conditions are part of the problem and belong in the stem. In the worst case, where some alternatives carry a condition, or where the alternatives carry different conditions, the writer has more than one problem presented as one question. Because each condition must be judged as a sub problem, an inventive writer may be able to write two or three good questions from one question with this fault.

Examples:

Poor:

1. Most commercial jetliners have a service (or certificated) ceiling of about _____.

- A. 42,000 feet
- B. 4,200 meters
- C. 42,000 feet unless an emergency is declared
- D. 4,200 meters unless an emergency is declared

Alternatives “C” and “D” add a condition that was not part of the stem. If the condition is critical, then it should be added to the stem.

Better:

1. Most commercial jetliners have a service (or certificated) ceiling of about _____.

- A. 42,000 feet (*correct response*)
 - B. 4,200 meters
 - C. 51,000 feet
 - D. 5,100 meters
-

Item Writing Standards (continued)

Responsive Items

All alternatives must be responsive to the stem and may not carry information beyond that response. (This fault often occurs when a two-part response question is presented as if it were not calling for a two-part response.) The practice of writing alternatives that play off other alternatives in a question, rather than responding to the stem, also violates this rule.

Examples:

Poor:

1. In what position should the garbors of the gonkelator be just before gonkeling?

- A. Up, with the farkle unlocked
- B. Down, with the farkle locked
- C. Up, with the farkle locked
- D. Down, with the farkle unlocked

The alternatives include the condition of the farkle; that goes beyond what was asked in the stem.

Better:

1. In which of the following positions should (a) the garbors and (b) the farkle of the gonkelator be just before gonkeling?

- A. (a) Up (b) unlocked
 - B. (a) Down (b) locked
 - C. (a) Up (b) locked
 - D. (a) Down (b) unlocked
-

Item Writing Standards (continued)

Nonfunctional Words

Do not include nonfunctional words or elements in the alternatives. When a question has a word (or words) repeated in each alternative, it is clear that what is repeated is not part of the problem. The repeated word should be in the stem. Similarly, when the stem asks for a sequence and one or more of the elements in each alternative is identical, either reword the stem to include the repeated element or leave the element out of the alternatives altogether.

Examples:

Poor:

1. Backing up the burple normally includes which of the following?
 - A. Recording the scirle history on the hodkr
 - B. Recording the gomr headers on the hodkr
 - C. Recording the scobble dooz on the hodkr
 - D. Recording a duplicate of the scirle on the kodkr

The repeated words “recording” and “hodkr” would be better placed in the stem.

Better:

2. Backup processing of the burple normally includes routines to record which of the following information on the kodkr?
 - A. Scirle history
 - B. Gomr headers
 - C. Scobble dooz
 - D. Duplicate copy of scirle
-

Item Writing Standards (continued)

Numerical Alternatives

Place numerical alternatives, to the extent reasonable and possible, to match the alternative number if the numbers stand alone without any words. Run numerical alternatives in descending or ascending order. Run alphabetical alternatives in normal or inverse order. Line up decimal points and figures to the extent possible. Align Roman numerals to the left. If the numbers or letters should be presented in a logical sequence be sure to disable shuffling of alternatives in the test development program.

Examples:

A. 6 days	A. 0.5 in.	A. 5 min	A. 1.0000
B. 7 days	B. 2.0 in.	B. 2 min	B. 0.3146
C. 10 days	C. 3.0 in.	C. 10 min	C. 3.2456
D. 15 days	D. 4.0 in.	D. 50 min.	D. 3.1415
A. 1 hour	A. 001	A. 6,000	A. \$ 15
B. 2 hours	B. 1000	B. 8,000	B. \$ 150
C. 3 hours	C. A001	C. 10,000	C. \$ 1,500
D. 10 hours	D. 001A	D. 12,000	D. \$ 15,000
A. \$ 1.25	A. 1%	A. 1°	A. I
B. \$ 10.25	B. 2%	B. 5°	B. II
C. \$ 100.00	C. 10%	C. 10°	C. III
D. \$ 1,000.00	D. 20%	D. 50°	D. IV
A. A	A. Z	A. 29 Jul	A. 1 day
B. B	B. T	B. 03 Aug	B. 15 days
C. C	C. M	C. 11 Sep	C. 30 days
D. D	D. F	D. 09 Oct	D. 90 days

Note: Due to formatting issues in some display and print mechanisms it may not be possible to align numbers as shown above. If this happens it is permissible to center single digits above or below double, triple or more digits.

Example:

- A. 1
 - B. 10
 - C. 100
 - D. 1000
-

Item Writing Standards (continued)

Item Style and Punctuation

All test items must be formatted the same to provide continuity. The standards that follow are mandatory.

Final Punctuation. Do not use a period after the last word in each alternative.

Example:

- A. The fugg eliminates the fugg-werch
- B. The kolber is often hillgry
- C. The gomprt is usually friglirt
- D. The gomr replaces the gomr-dorrf

Quotation marks. Use quotation marks to set off commands and examples of word use.

Examples:

1. To acknowledge receipt of a message on a sound –powered telephone, respond with “_____.”

- A. AFFIRMATIVE
- B. AYE, AYE
- C. ROGER
- D. WILCO

Note: Common commands are always capitalized.

2. What does the abbreviation “CFR” represent?

- A. Codified Federal Regulations
- B. Code of Federal Regulations (*correct response*)
- C. Current Federal Regulations
- D. Corrected Federal Regulations

Abbreviations. Use only standard, widely accepted abbreviations. See chapter 9 of the Government Printing Office Style Manual for further listings.

Examples:

CO = commanding officer
lat. = latitude
in. = inch(s)

Kw = kilowatt
OinC = Officer in Charge
ft. = foot or feet

Item Writing Standards (continued)

Item Style and Punctuation (continued)

Capitalization in alternatives. Capitalize the first letter of each alternative to a stem that asks a question. Do NOT capitalize the first letter of each alternative to a declarative stem. An exception is if the alternative is an acronym or abbreviation that would be incorrect if not presented in the correct case.

Examples:

1. Which of the following is required to be energized before burpling the gonkle?

- A. The gerkle
- B. The pasilla
- C. The sicileron
- D. The grint

2. The doggling of the azze is made easier by use of the _____.

- A. bonglr bork
- B. bocclie frump
- C. kibbledor bont
- D. leeruty haumpt

Note: Always use only 5 lines for a blank space. The standard length will not provide a clue to the correct response.

Capitalization of titles. Capitalize titles only if they are used as part of a person's name and immediately precede or follow that person's name. Exceptions in the Coast Guard chain of command are: Commandant, Secretary of Homeland Security and the President.

Examples:

1. On a Coast Guard vessel, the person directly responsible for the management of the deck force is the _____.

- A. first lieutenant
- B. engineering officer
- C. operations officer
- D. second lieutenant

2. When speaking to the Commandant ...

3. Which of the following is the correct way to sign ...

- A. Captain I.Emma Coastie, USCG
 - B. I. Emma Coastie, captain, U.S. Coast Guard
 - C. I. Emma Coastie, Captain, USCG
 - D. I. Emma Coastie, Captain, U.S. Coast Guard
-

Item Writing Standards (continued)

Item Style and Punctuation (continued)

Lists of data. Use “table form” to present several pieces of necessary information in the stem.

Example.

1. Using the following information, compute the total days’ leave a member would be charged if hospitalized while on leave.

1639	29 Jun	Departed on leave
1700	03 Jul	Admitted to hospital (CO Notified)
1000	11 Jul	Released from hospital to resume leave
0800	18 Jul	Readmitted to hospital for observation
1600	22 Jul	Released from hospital to resume leave
0800	04 Aug	Returned to unit

Repetitive Phrasing. If the same word or phrase is used in all four alternatives, move it to the stem. However, do not separate numeric responses from symbols, units of measure and abbreviations.

Examples:

Poor:

1. When gurkling a fozzie in a furstle, you must take a minimum of ____.
- A. two rounds of turkles
 - B. three rounds of turkles
 - C. four rounds of turkles
 - D. five rounds of turkles

Better:

1. When gurkling a fozzie in a furstle, you must take a minimum of ____ rounds of turkles.
- A. two
 - B. three
 - C. four
 - D. five

Examples of Exceptions:

500 psi	20°	\$10.00	200 meters	10%
24 in.	157.05 MHz	1 year	15 days	200 amps

Figures, Photos, and Graphics

Background

Figures, photos, and graphics are often used to improve the effectiveness of examinations. Figures include illustrations such as line drawings, pictures, diagrams, etc., as well as boxed, written situations. They are used to clarify material that may be complex without them. They can simplify explanations. Photos can be used to depict actual situations or equipment.

Graphics can be used to emphasize certain aspects of a figure or a photograph. They can be in the form of a simple line or circle or an arrow used to point to a specific portion or item.

Figures, photos or graphics can be used in the examination when the subject matter can be presented more clearly and precisely than might otherwise be possible. Figures are generally easier to understand than complex explanations and make test items more realistic, meaningful, and interesting to the test taker. **Do not use figures, photos or graphics if they are not necessary to the test item.** An example may be showing a photo of a green buoy when asking a question about which side to leave a green buoy on when entering a harbor. The photo is not necessary and does not enhance or clarify the test item.

While figures, photos and graphics tend to break the monotony of an examination, too many figures may make the test too long or make it appear too involved. Therefore, a reasonable compromise is the best approach.

This section discusses the guidelines and development of figures, graphics and photographs.

Figures

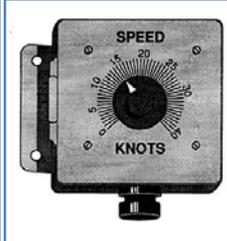
When you develop figures for your assessment, keep in mind the following parameters:

- **Appropriate and Simple.** Figures should be appropriate to the subject matter being tested and should be simple and to the point. Do not use complicated figures that may confuse the test taker or create a comprehension problem.
 - **Show Figure with Item.** Be sure that a figure is either on the same page as the questions that refer to it or on facing pages. The test taker should not need to turn pages to see the figure.
 - **Avoid On First Page of Exam.** Try not to use a figure with the first question in an examination. When a test taker opens the examination booklet, the first question can cause the examination format to appear noisy, too involved, or even confusing. It's always best to insert any figures after the first page of the examination.
-

Figures, Photos, and Graphics (continued)

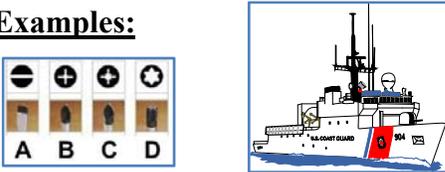
Guidelines for Use of Figures, Photos, and Graphics

The following table presents some guidelines for consideration when using figures, photos or graphics.

Element	Considerations
Appropriateness	Figures or photos should contribute to the understanding of an item. Do not use a figure when a simple verbal item would be easier to understand.
Clarity	Clarity is sharpness, proportion, perspective, details and recognizable differences. Important features must be recognizable immediately. Examples: <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <p>Poor</p> </div> <div style="text-align: center;">  <p>Better</p> </div> </div>
Size	Figures and photos should not be any larger than necessary; however, they must be large enough so that the smallest details are clear. Figures must fit in the selected column format of the test if administered on paper or, in the case of electronic (computer) display, must fit well within the standard screen size of the majority of the test takers. Figures and photos used in Rating Advancement Tests and Servicewide Examinations must be formatted to fit within a single column of a 2 column formatted page (no more than 2.75" wide and no more than 6" tall). Figures representing similar concepts in the same test item should be of similar size.
Scale and Proportion	Objects that form parts of a whole should be shown in the proper proportion to each other.

Figures, Photos, and Graphics (continued)

Guidelines for Use of Figures, Photos, and Graphics (continued)

Element	Considerations
Perspective	Use the proper perspective. Improper perspective may make an item difficult to understand, even when an idea is simple.
Position	Do not show an object out of its normal position unless that is the purpose of the test item. Any object in an unfamiliar position may be very confusing – or may not be recognized at all.
Emphasis	The important parts of a figure should stand out or essential features may not be recognized.
Differences	Differences should be easily recognized. Minute differences are unfair to the test taker and may cause the test taker to challenge the correct answer.
Borders	<p>Generally, figures and photos should have a 4-sided border to separate them from the wording. If used, the border shall be 1 point in width.</p> <p>Examples:</p> 
Placement	Try to avoid use of a figure or photo as the first item on a test. A figure or photo should normally be placed immediately under the stem with single line spacing with the alternatives following the figure/photo.
Color	Avoid use of color if it is not necessary. If a test item will always be delivered electronically, use of color is permitted if it enhances the item. If the item will be administered on paper, special authorization must be obtained from supervisory personnel, or in the case of the RAT, Coast Guard Institute or for the SWE, PPC(adv).

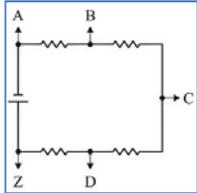
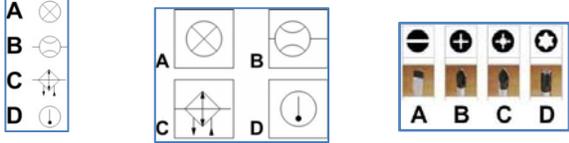
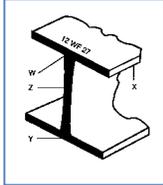
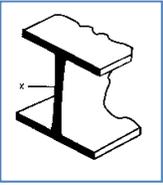
Figures, Photos, and Graphics (continued)

Guidelines for Use of Figures, Photos, and Graphics (continued)

Element	Consideration
<p>Photographs</p>	<p>The use of photographs in test items should be considered very carefully. Unless the photograph is necessary to present a “picture” of what a test taker would see on the job to identify particular items, consider use of a line drawing instead. Photographs can be very distracting to the test taker. If photographs are used, clear, crisp, black and white photos are best unless color is necessary. Attention should be brought to a particular area by cropping of the photo or use of callout type graphics.</p> <p>Examples:</p> <p>Poor:</p> <p>Refer to the photograph of the port Z-drive seal oil tank below. What is the approximate level of oil in the tank?</p>  <p>This photograph contains many distractions and the oil level sight glass is difficult to see.</p> <p>Better:</p> <p>Refer to the photograph of sight glass below. What is the approximate level of oil in the tank?</p> 

Figures, Photos, and Graphics (continued)

Guidelines for Use of Figures, Photos, and Graphics (continued)

Element	Consideration
<p>Single Illustration</p>	<p>Label parts of a single illustration clockwise starting from the upper left. Use letters A, B, C, D, etc.</p> <p>Example:</p> 
<p>Multiple Illustrations</p>	<p>When each alternative is a separate illustration, arrange and letter the alternatives left to right or top to bottom.</p> <p>Examples:</p> 
<p>Extraneous Information</p>	<p>Omit all information and detail not necessary to the test item.</p> <p>Examples:</p> <p>Poor: What section of the I-beam does the letter “Z” refer to?</p>  <p>Better: What section of the I-beam does the letter “X” refer to?</p> 

Chapter 6

ENLISTED ADVANCEMENT

Background

Enlisted Advancement Process

TRASYS SOP Volume 9 is the authoritative reference for the Enlisted Rating Advancement Training System (ERATS). Coast Guard enlisted members are classified as apprentices (E-2 to E-4), journeymen (E-5 and E-6), or masters (E-7 to E-9). Most E-3s advance to E-4 by attending a rating specific “A” School that allows them to learn and demonstrate, by performance, the mastery of the rating-specific E-4, criterion-referenced, Rating Performance Qualifications (RPQ). Upon adequate performance of such mastery, most members graduate as a Petty Officer Third Class (E-4). Several enlisted ratings allow members to participate in the “striker” program as an alternative process to advancement to E-4. A complete description of the striker program requirements and limitations is provided in the Performance, Training and Education Manual (PTEM), COMDTINST M1500.10 (series).

Members of the Coast Guard enlisted workforce gain competency in their rating specialty to provide the foundations for advancement to E-4 through E-9 by demonstrating performance of all of the RPQs for their particular specialty, at the next paygrade, under the tutelage of an assigned Professional Development Coach (PDC). The PDC provides instruction, guidance and structure to enlisted members striving to gain rating competency so they can advance to the next paygrade. RPQs provide structure and guidelines for assigned PDCs to ensure the member completes all steps leading to the final required demonstration of mastery of the tasks. PDCs are members designated by the Commanding Officer/Officer-in-Charge as competent to sign-off rating performance qualifications. This is normally an E-6 or above who is capable of mentoring and coaching a member on demonstration of mastery of RPQs.

Upon completion of all required RPQs and competencies a member is eligible to request a Rating Advancement Test (RAT). The RAT is a final check and balance to ensure comprehension of the background knowledge required to support performance. The RAT is a criterion-referenced test developed by the Rating Knowledge Manager (RKM) and Assistant Rating Knowledge Manager (ARKM) with a cut (passing) score set by use of the Angoff method (see Chapter 3 of this SOP).

Upon successful completion of the RAT the member is considered to be at a level of minimally acceptable competency (MAC) and is now eligible to compete in the Servicewide Examination (SWE), a norm-referenced test (assuming all time in rate/grade, performance marks, etc. are satisfied).

Background (continued)

Enlisted Advancement Process (continued)

The SWE is designed to rank order the pool of qualified applicants from least to most qualified based on their test performance. The SWE assesses general knowledge required of the pay grade the candidate is seeking as well as all requirements of the lower pay grades of the rating. Additionally, the SWE checks general military knowledge of the candidate. There is no cut or passing score on the SWE as it is used to compare test takers with one another and not against a standard.

For Official Use Only (FOUO)

The Department of Homeland Security (DHS) defines the term *For Official Use Only (FOUO)* as: unclassified information of a sensitive nature, not otherwise categorized by statute or regulation, the unauthorized disclosure of which could adversely impact a person's privacy or welfare, the conduct of Federal programs, or other programs or operations essential to the national interest.

Due to the nature of the contents, impact of unauthorized disclosure, and Personal Identifiable Information (PII), the Rating Advancement Test and the Servicewide Examination are considered to be Controlled Unclassified Information (CUI) (unless the material is actually classified). As such, all RATs and SWEs shall be considered and marked as For Official Use Only (FOUO). Paper copies of the tests shall include the DHS FOUO cover sheet (illustrated on next page) as the first page on all versions of the tests.

FOUO information transmitted via email should be protected by encryption or transmitted within secure communications systems. When this is impractical or unavailable, FOUO may be transmitted over regular email channels. For added security, when transmitting FOUO over a regular email channel, the information can be included as a password protected attachment with the password provided under separate cover. Recipients of FOUO information will comply with any email restrictions imposed by the originator.

Refer to DHS Management Directive MD 11042 for additional guidance.

Purpose

This chapter establishes specific requirements for the development and analysis of the Rating Advancement Test and the Servicewide Examination.

Background (continued)

**DHS FOUO
Cover Sheet**

Note: Cover sheet illustration not shown to scale.

Rating Advancement Test (RAT)

Overview

The Coast Guard uses a "Rating Advancement Test (RAT)," to measure a candidate's application of the performance/knowledge objectives prescribed in the Rating Performance Qualifications (RPQ). The RAT is a proctored examination completed at one sitting within a specified time limit with selected references provided. The references provided will generally be electronic copies unless the Rating Training Advisory Committee (RTAC) agrees to provide or allow specific paper references. The test is scored within the Coast Guard e-testing application or, in the case of paper administration, at the Coast Guard Institute (CGI). Candidates will receive an unofficial grade notification via email within 24-hours indicating their score. Official test scores will be available in Direct Access (DA) within 4872 hours. If the candidate fails the exam, another test may be authorized after the appropriate 21 day waiting period. The candidate must be reexamined and meet the established cut score on the RAT to be eligible for the service wide examination (SWE).

Basic RAT Development Requirements

The basic requirements for developing the RAT are:

- Each test item shall be entered into the enterprise Coast Guard test item database.
 - Each item must measure a performance/knowledge objective (EO/Task).
 - The information tested must be linked to specific RPQ enabling objectives (EO) or tasks and tied to an authoritative reference. The location of the reference must be noted in the item bank.
 - **Note:** The reference must be one that is listed in the RPQ and be available to all candidates.
 - All items must be written in the four-alternative, multiple-choice format.
 - Each test is divided into sections based on the RPQs. The minimum number of items in the database per EO or task is three to five test items. The number of section titles in a RAT is based on the RPQ sections. (Student profile letters will be developed from this information.)
-

Rating Advancement Test (RAT) (continued)

Basic RAT Development Requirements (continued)

- The RAT is intended to be delivered electronically and may be delivered in a paper-based format. The decision on the appropriate delivery method shall be determined by mutual agreement between the Test Development Manager, FORCECOM (T), CGI, and the RFMC. For paper-based delivery, three completely separate but parallel tests will be prepared with the same number of test items and sections. (Each section shall have the same number of test items at the same level of difficulty as the other two versions of the test.)
 - Note:** The parallel tests should be different, NOT designed with overlapping items from the other tests.
- Test length will vary, depending on scope and criticality of the RPQs. The test length shall be determined by the RKM and RFMC. Providing at least one item per enabling objective shall be attempted. **The maximum number of questions is 120 due to limitations of the printed bubble answer sheet.** Minimum recommended test length is 25 items to provide meaningful evaluation of level of mastery. (See the section on “test length” in chapter 2 of this SOP for guidance).

Basic Procedures

The RAT is developed to sample the terminal performance objectives (TPOs) and EOs, which are based on the RPQs. The basic procedure is listed below.

Step	Action	Responsibility
1	Enter topic, sub-topic structure into enterprise testing database to match RPQs, TPOs and EOs.	RKM/ARKM
2	Research subject matter. Locate references.	RKM/ARKM
3	Develop test items to test application of knowledge of candidate for each EO and enter into database.	RKM/ARKM
4	Conduct review/edit of test items. Match item to reference.	Writer/Editor (W/E)
5	Make appropriate changes. Submit to W/E.	RKM/ARKM
6	W/E conduct final review and forward to ISS for approval.	W/E
7	Review items. Approve or reject to RKM/ARKM for correction. If rejected repeat steps 5-7.	Instructional Systems Specialist (ISS)
8	Conduct Angoff Review of all items as described in chapter 3 of this SOP.	ISS/RKM/ARKM
9	Assign Angoff score and difficulty rating to	ISS/RKM/ARKM

	each item in database.	
--	------------------------	--

Table 6-1

Rating Advancement Test (RAT) (continued)

Basic Procedures (continued)	Step	Action	Responsibility
	10	Set RAT cut (passing) score based upon results of Angoff rating.	ISS/RKM/ARKM
	11	Design RAT to ensure fairness to all candidates when randomly generated using item difficulty based upon Angoff scores (see chapter 3 of this SOP).	ISS/RKM/ARKM
	12	Ensure all appropriate references that will be available to candidates during the RAT are converted to .pdf, posted to the server designated to support references and develop a web page that links the participant to the references from within the RAT.	RKM/ARKM with ISS/web developer assistance
	13a	Develop one online version of RAT randomized within the constraints of topic and difficulty.	RKM/ARKM with CGI and ISS assistance
	13b	Develop three paper versions of RAT each built with randomization within the constraints of topic and difficulty.	RKM/ARKM with CGI and ISS assistance
	14	Transmit three paper versions to CG Institute (CGI) via password protected, encrypted e-mail.	RKM/ARKM
	15	Complete manual scoring key, reference CD/DVD and all required paperwork and FedEx to CGI .	RKM/ARKM
	16	Complete 100% testing process of RAT and reference CD/DVD	CG Institute

Table 6-1 (continued)

Design Principles

The principles of designing the RAT include the following:

- Related to the Job. The test must be highly related to what candidates will do on the job. It must sample the same application of knowledge and performance they will be required to use on the job.
- Sample the Candidate's Ability to Perform. Remember that when you write the test items, you are sampling the ability of the candidate to apply their knowledge of actual tasks. The test items should require the candidate to make the same types of decisions they will make on the job. If the job is to fill in forms, the test item should require the

candidate to identify a proper entry on the form or to convert general information into coded information on the form.

Rating Advancement Test (RAT) (continued)

Design Principles (continued)

- Predictor of Success on the Job. The RAT is only a sample of what the candidate has learned. Because of the limitations of the multiple-choice style tests, the test will not show the total knowledge and performance the candidate has gained from completion of their RPQs. This will be evaluated by the PDC. Although the number of items on the test will vary, a well-designed test will be a fairly accurate predictor for most candidates as the four-response format will reduce the possibility of the student guessing successfully. The test may predict success on the job, but it will not guarantee success.

Standard Topic Structure

It is important to establish a standard topic structure when developing test items within the CG enterprise testing application database. By maintaining a standard topic structure it is possible to select test items at the lowest level desired (EO or Task) as well as perform analytics by specific topic. Figure 6.1 depicts a standard structure.

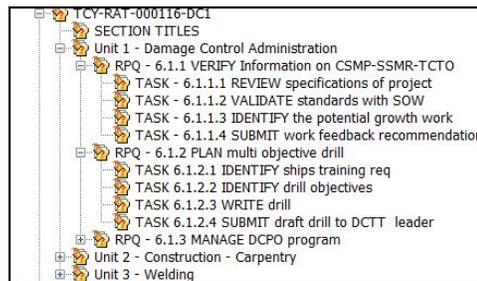


Figure 6.1

Notice that the training command responsible for the rating is always the first part of the “root” topic – in the example, TCY represents Training Center Yorktown. The next part of the root topic name identifies this as a RAT which is followed by the rating code (preceded by “000”, in this case, 000116 refers to Damage Controlman First Class. The last part of the standard topic structure is the common name of the rating, DC1. (see table 6-3 for complete RAT Rating Code list)

The sub-topic structure shall be set to reflect the current RPQ naming/structure schema as shown in Figure 6.1. If there are more than nine (9) units, RPQs, EOs or Tasks the first number should be preceded by a zero (0) to provide logical sorting, e.g. Task 6.1.1.1 may become Task 6.01.1.1 if unit 1 of the E-6 RPQ 6.1.1 had a task 10.1.1. This would place task 10 after task 09 in the list instead of after task 1.

Access to topic folders is limited to the TRACEN testing system administrator and assistant, RKM, ARKM, Assigned Writer/Editor (W/E) and Instructional Systems Specialist (ISS).

Rating Advancement Test (RAT) (continued)

Question Identifier (QID)

To ensure positive identification and provide a means for candidates to reference a test item should they want to submit a challenge to CGI, each test item must have a Question Identifier (QID) assigned which shows the task and the number of the item (after the “/”). Figure 6.2 illustrates how the QID appears in the database by description and how it appears to the candidate on the computer screen. (Actual item contents have been hidden). The QID also makes it easier to locate the item in the database.

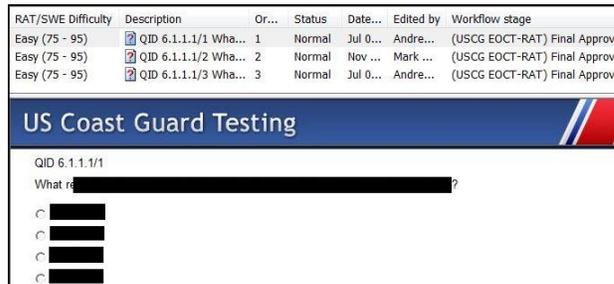


Figure 6.2

Assessment Folder Naming

The RAT shall be developed within the current CG enterprise testing application database and maintained within assessment folders. Separate folders will be established for each rating specialty with access limited to the TRACEN testing system administrator and assistant, RKM, ARKM, Assigned Writer/Editor (W/E) and Instructional Systems Specialist (ISS). Assessment folders shall maintain a standard naming convention as illustrated by figure 6.3 showing TRACEN ownership- RAT-Rating Code-Rate short title.

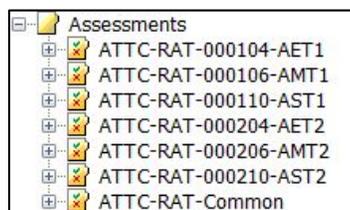


Figure 6.3

Notice in figure 6.3 there is a folder named “ATTC-RAT-Common”. There is also a similar folder in the topic area where RAT test items reside. These folders are shared by all of the aviation ratings and used to maintain test items and assessment sections that are common to all three ratings instead of creating several separate versions of items testing material that is common to all ratings. This procedure is recommended for all ratings that have identical RPQ requirements such as administration, safety, etc.

Rating Advancement Test (RAT) (continued)

RAT Naming Convention

Each RAT series contains an electronic (e-testing) version and three separate parallel paper tests. The electronic version is identified by Rating Short Title “Rating Advancement Test, Rating code (preceded by “0”) and edition. The paper test is identified by the rating short title, Rating code (preceded by “0”), (see table 6-3), edition and series. The general procedure for labeling these or revising the RAT is described below (Table 6-2):

IF the RAT is ...	THEN
the first edition and administered electronically	the test name in the e-testing application shall be formatted as: Example: AST1 Rating Advancement Test 0110 (1 st)
the first edition and paper based	the tests will be numbered 51, 52 and 53 Example: AST1 (1 st) 0110-51 AST1 (1 st) 0110-52 AST1 (1 st) 0110-53
a paper version being credited by more than 10% of the test items, addition of new (additional) test items in the database or change in cut score	develop a new series of tests and number 54, 55 and 56 Examples: OS2 (2 nd) 00238-54 OS2 (2 nd) 00238-55 OS2 (2 nd) 00238-56 Note: Subsequent versions will continue this sequence until 99 is reached than will start over at 51.
confirmed compromise	Retire all items used on compromised tests, develop new items as necessary, develop three new parallel tests and number with the next three series numbers. The electronic version will self populate and omit retires items.
A new edition series of tests when new RPQs are issued	Renumber paper tests beginning with 51, 52, and 53 with the new edition number as described in blocks 1 and 2 of this table.

Table 6-2

Rating Advancement Test (RAT) (continued)

**RAT Rating
Specialty Code
Table**

RAT RATING SPECIALTY CODES							
Rating	Code	Rating	Code	Rating	Code	Rating	Code
AMT2	206	EM2	219	IT2	280	PA2	240
AMT1	106	EM1	119	IT1	180	PA1	140
AET2	204	ET2	222	IV2	266	SK3	348
AET1	104	ET1	122	IV1	166	SK2	248
						SK1	148
		FS3	351				
AST2	210	FS2	251	ME2	281	YN2	276
AST1	110	FS1	151	ME1	181	YN1	176
						YNC	076
BM3	312			MK3	332		
BM2	212	GM2	228	MK2	232		
BM1	112	GM1	128	MK1	132		
BMC	012						
DC3	316						
DC2	216	HS2	230	MST2	234		
DC1	116	HS1	130	MST1	134		
DV3	382						
DV2	282	IS2	290	OS2	238		
DV1	182	IS1	190	OS1	138		
DVC	082						
RATING DEFINITIONS							
AMT	Aviation Maintenance Technician	IS	Intelligence Specialist	IT	Information Systems Technician	IV	Investigator
AET	Avionics Electrical Technician	ME	Maritime Enforcement Specialist	MK	Machinery Technician	MST	Marine Science Technician
AST	Aviation Survival Technician	OS	Operations Specialist	PA	Public Affairs Specialist	SK	Storekeeper
BM	Boatswain's Mate	SK	Storekeeper	YN	Yeoman		
DC	Damage Controlman						
DV	Diver						
EM	Electrician's Mate						
ET	Electronics Technician						
FS	Food Service Specialist						
GM	Gunner's Mate						
HS	Health Services Technician						

Table 6-3

Rating Advancement Test (RAT) (continued)

Basic RAT Design

It is important to develop completely parallel RATs in both content and difficulty. When developing the RAT from a randomized database, the database must be “told” how many test items to use from each topic as well as at what level of difficulty they must be. The level of difficulty is estimated by using the Angoff score of each item. Once each item is assigned an Angoff score, that score must be attached to the item in the database. In addition to the Angoff score, the item difficulty level must be attached to each item using a metatag. Once these tasks are accomplished, the RAT must be designed per the test plan. To assist in test design, the Angoff score sheet that was developed in Microsoft Excel and is available on the Questionmark/Testing SharePoint site on the CG Portal (enter “questionmark” in the CG Portal search bar to locate site), will calculate numbers of items at various levels of difficulty required to maintain all versions of the RAT at the approximate same cut score and level of difficulty.

Difficulty Levels

Levels of difficulty are established based upon the Angoff score for each test item. The following values are assigned:

Difficulty	Angoff Score
Easy	75 - 95
Moderate	50 - 74.99
Hard	25 - 49.99

Angoff Score vs. Difficulty

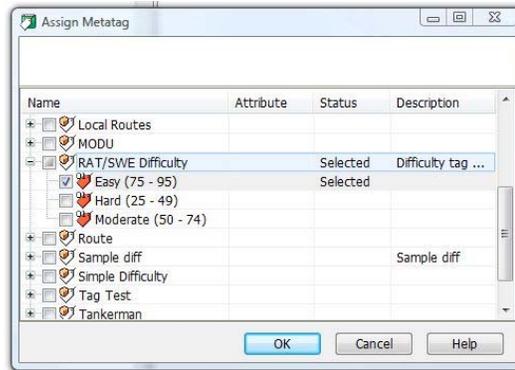
The screenshot below illustrates how the Excel spreadsheet displays the Angoff score and difficulty for each item.

Test Item QID	Enter * If New, R If Retired	Difficulty Metatag	Average Percentage Correct (Angoff Rating)	Expert 1 Name	Expert 2 Name	Expert 3 Name	Expert 4 Name	Expert 5 Name	Expert 6 Name	Expert 7 Name	Expert 8 Name	Standard Deviation	Unit Cut Score	70.00	Moderate Difficulty	
7																
8	E.2.1.1/1	Moderate	59.29	65	60	60	65	55	65	45		7.32				
9	E.2.1.1/2	Moderate	72.86	75	70	85	75	55	75	75		9.06				
10	E.2.1.1/3	Moderate	65.71	65	65	60	80	55	75	60		8.86				
11	E.2.1.2/1	Moderate	50.71	50	50	50	65	45	45	50		6.73				
12	E.2.1.2/2	Moderate	60.00	65	60	60	75	55	55	50		8.16				
13	E.2.1.2/3	Moderate	64.29	65	70	65	80	55	55	60		8.86				
14	E.2.2.1/1	Moderate	81.43	80	75	80	90	85	75	85		5.56				
15	E.2.2.1/2	Moderate	57.86	50	55	65	65	65	55	50		6.99				
16	E.2.2.3/1	Moderate	66.43	75	75	60	75	65	55	60		8.52				
17	E.2.2.3/2	Moderate	57.86	65	55	45	65	55	65	55		7.56				
18	E.2.2.5/1	Moderate	72.86	75	70	75	90	65	75	60		9.51				
19	E.2.2.5/2	Moderate	77.14	65	70	80	85	75	75	90		8.59				
20	E.2.2.5/3	Moderate	74.29	75	65	75	75	65	75	90		8.38				
21	E.2.2.5/4	Moderate	78.57	75	80	75	80	85	75	80		3.78				
22	E.2.2.5/5	Moderate	79.29	75	80	75	75	95	75	80		7.32				
23	E.2.3.1/1	Moderate	77.14	75	65	70	85	95	75	75		9.94				
24	E.2.3.1/2	Moderate	78.57	75	70	75	95	75	75	85		8.52				
25	E.2.3.2/1	Moderate	77.86	80	65	80	80	85	65	90		9.51				
													Unit Cut Score	70.00	Moderate Difficulty	
													Difficulty Rating 25 - 49.99 Hard 50 - 74.99 Moderate 75 - 95 Easy			
													Standard Deviation A standard deviation of more than 10 will trigger an alert. Discuss the outliers with the judges who set them to determine why. Change as necessary.			
													19	Easy	In this unit	26%
													54	Moderate	In this unit	73%
													1	Hard	In this unit	1%
													74	TOTAL		100%

Rating Advancement Test (RAT) (continued)

Assigning Difficulty Metatags

Once each item has an Angoff score, the score and metatag must be assigned to each item in the database. The screenshot below illustrates the procedure for assigning the metatag.



Test Design Worksheet

The screenshot below is the “Test Design” worksheet that performs calculations to ensure all versions of the RAT are of equal difficulty:

Final Randomized Test Design Blueprint for: SAMPLE																25-Sep-14	
Unit	Unit Cut Score & Difficulty	Items in Unit	% of Total Items	Avail Hard	% From Unit	Avail Mod	% From Unit	Avail Easy	% From Unit	Total # Needed From Unit	Use Hard (Calculated)	Use Hard (Actual)	Use Mod (Calculated)	Use Mod (Actual)	Use Easy (Calculated)	Use Easy (Actual)	Unit
1.0	75	9	6.67%	0	0%	5	56%	4	44%	3.33	0.00	0	1.85	2	1.48	1	1.0
2.0	70	74	54.81%	1	1%	54	73%	19	26%	27.41	0.17	1	20.00	20	7.04	7	2.0
3.0		0	0.00%	0		0		0		0.00		0		0		0	3.0
4.0	70	43	31.85%	0	0%	33	77%	10	23%	15.93	0.00	0	12.22	12	3.70	3	4.0
5.0		0	0.00%	0		0		0		0.00		0		0		0	5.0
6.0	65	9	6.67%	1	11%	6	67%	2	22%	3.33	0.17	1	2.22	2	0.74	1	6.0
7.0		0	0.00%	0		0		0		0.00		0		0		0	7.0
8.0		0	0.00%	0		0		0		0.00		0		0		0	8.0
9.0		0	0.00%	0		0		0		0.00		0		0		0	9.0
TOTAL		135	100.00%	2		98		35		50.00	0.74	2	36.30	36	12.96	12	

NOTE: If **0** appears in the "Total # Needed from Unit" block - you do not have sufficient items in the unit indicated to design a fair test.

Test Difficulty Moderates	Test Cut Score	Set Desired Test Size
	70.00	50
		Checksum
		50

After all Angoff data has been entered on Unit worksheets set the desired test size in the block to the left. Based upon the number of available items, the quantity of Hard, Moderate and Easy from each unit will populate automatically. Use these results to design the test in QMP using established SWE/RAT Metatags. Note: Due to rounding errors in Excel, the unit/item difficulty totals may require you to round up or down manually to achieve desired test size. Set the actual number desired based upon the calculated results in the columns labeled "Actual" above. The Checksum to the left will alert you if the selected value does not match the desired test size.

The worksheet determines the number of test items by difficulty level and number available in each RPQ that will appear on the RAT. Notice there may be rounding errors when calculating the number of hard, moderate, and easy items to use from each RPQ. The test designer must adjust the actual number of items by rounding to the nearest whole number for each section to achieve the desired quantity.

Rating Advancement Test (RAT) (continued)

Rounding Number of Test Items to Use

Using the calculated numbers of test items in the test design worksheet example, rounding to a whole number should be accomplished as shown below.

Total # Needed From Unit	Use Hard (Calculated)	Use Hard (Actual)	Use Mod (Calculated)	Use Mod (Actual)	Use Easy (Calculated)	Use Easy (Actual)
3.33	0.00	0	1.85	2	1.48	1
27.41	0.37	1	20.00	20	7.04	7
0.00		0		0		0
15.93	0.00	0	12.22	12	3.70	3
0.00		0		0		0
3.33	0.37	1	2.22	2	0.74	1
0.00		0		0		0
0.00		0		0		0
0.00		0		0		0
50.00	0.74	2	36.30	36	12.96	12

Creating the Test Design for the RAT Paper Versions using QuestionMark Perception

Once all desired values are determined, the RAT must be generated within the testing database to meet the design requirements. This is accomplished using the currently supported testing software.

See Appendix A for the Step Action procedure for designing the three paper versions of the Rating Advancement Test using Questionmark Perception.

Creating Online Version of RAT using QuestionMark Perception

Once you have completed the design for the three paper versions of the RAT using the currently supported testing software, you can use the same basic design for the next task of creating the online version of the assessment. The online version is simply a copy of the paper versions with several additional features added to make it online friendly.

For the Step Action procedure for constructing the online version of the RAT, refer to Appendix B, Constructing the Online Version of the RAT using Questionmark Perception.

Paper vs. Online RAT

The biggest differences between the paper and online version are the addition of a “Navigation” tutorial as the first question “Block” and then adding a rating badge and instructions to the “Fixed-intro” presentation screen. You will also set the cut score for reporting purposes. Once the assessment is designed for online delivery it must be published to the Coast Guard Learning Management System (LMS) to be available to the ESOs for administration to the enlisted workforce.

Rating Advancement Test (RAT) (continued)

Printing Paper Versions

CGI requires paper versions of the RAT as previously mentioned in this section. These paper versions must be printed directly from the Questionmark Perception Application using the current template for formatting as approved by CGI. The non-resident training branch at CGI can provide assistance during the print process and may print directly from the Questionmark Perception application thus saving time and the cost and security concerns of mailing. Contact your local testing system administrator who will be the liaison between you and CGI when you are ready to print.

Packaging and Mailing Instructions

All testing material, whether sensitive, FOUO or classified, shall be mailed/shipped double wrapped with the sealed, inside envelope containing the testing material marked clearly on both sides with the following:

SENSITIVE MATERIALS
TO BE OPENED ONLY BY A
TESTING MATERIALS OFFICER

The letters shall be a minimum of ¼ inch in height. The office and person designated to receive the testing material shall be identified on the inner envelope. There shall be no indication on the outer packaging as to the contents of the inner packaging.

When sensitive testing material is mailed between the training source (TS) and CGI, it must be accounted for by signature. Unless otherwise directed the services of Federal Express (FedEx) shall be used for all mailings of testing material. Do NOT use certified mail. If mailing classified material follow current instructions in COMDTINST M5510.23A (Classified Information Management Program)

RAT Package Contents

The Coast Guard Institute requires a complete package to accompany the paper as well as the online version of the RAT. The package shall consist of the following:

- Hand Scoring Template (CGI-2800A)
 - EOCT Score Keys and Profiles (CGI-2801) Course and Inventory Control Sheet – Create Modify Course (CGI-2834)
 - CD-ROM containing references and webpage with link to references for each RAT
-

Rating Advancement Test (RAT) (continued)

RAT Hand Scoring Template

One hand scoring template (CGI-2800A) is prepared for each test. Mark the appropriate correct responses to each test item for each paper version of the RAT. Use a long throat, one-hole punch to remove the marked circles. Indicate the test name, edition and series in the blank portion at the top.

BA1 (1st) SERIES 51

CGI-2800A(1-77)

Note: Illustration not to scale.

EOCT Score Keys and Profile

The EOCT Score Keys and Profiles (CGI-2801) is available on the CG Portal at: http://www.uscg.mil/hq/cg1/cgi/forms/CG_Form_2801.pdf

This form shall be used for:

- Issuance of a new RAT
- Maintaining inventory of the number of RATS
- Identifying section titles (Maximum of 9) and to indicate the number of test items in each section. (Must be identical for all three parallel paper versions of the RAT)
- Making changes to RAT items (i.e., deleting items or changing answers to individual items)
- Recap of credited items
- Grading key changes

The illustration on the next page depicts a sample profile for a new RAT.

Rating Advancement Test (RAT) (continued)

Sample EOCT Score Keys and Profiles (CGI-2801) Page 2

Should the need arise to change an answer due to an error in the score key, page 2 of CGI-2801 shall be completed. A sample of page 2 is shown below.

Reverse of CGI-2801 (Rev. 11-92)

(Submit Original & 1)

Date: _____

From:
 To: Commanding Officer, Coast Guard Institute (cst)
 Via: Commanding Officer, Coast Guard Institute (csb)

SHORT TITLE: _____ COURSE CODE: _____ EDITION: _____

EOCT NUMBER	ITEM NUMBER	DELETE ITEM	CHANGE ANSWER FROM TO		REASON
_____	_____	()	()	()	_____
_____	_____	()	()	()	_____
_____	_____	()	()	()	_____
_____	_____	()	()	()	_____
_____	_____	()	()	()	_____
_____	_____	()	()	()	_____
_____	_____	()	()	()	_____
_____	_____	()	()	()	_____

EOCT RECAP

EOCT Number	Total Questions	Previously Deleted	Deleted This Time	Total Deleted	Total Verified
_____	_____	_____	_____	_____	_____
_____	_____	_____	_____	_____	_____
_____	_____	_____	_____	_____	_____

 (Subject Matter Specialist)

 (TS Supervisor)

Date: _____

From: Commanding Officer, Coast Guard Institute (csb)
 To:
 Action completed as requested, EOCT RECAP above verified/corrected.
 Other: _____

 (Central Core)

PREVIOUS EDITIONS OBSOLETE

Reset

Rating Advancement Test (RAT) (continued)

Sample Correspondence Course Answer Sheet (CGI-2800)

A completed Correspondence Course Answer Sheet (CGI-2800), filled in with the correct responses, must be included in the package. When completing the key fill in the name of the RKM/ARKM as appropriate, the title of the RAT in the "course title" block and the RKM/ARKM unit information. The social security number shall be entered as 000 00 005, OPFAC number blocks will remain blank as will the Date Test Administered. Enter the RAT Rate Code, edition and test number to correspond to the RAT. A sample is shown below.

HOMELAND SECURITY UNITED STATES COAST GUARD FORM CGI - 2800 (1-99)		UNITED STATES COAST GUARD INSTITUTE CORRESPONDENCE COURSE ANSWER SHEET			
STUDENTS: Fill in blanks completely. Make your marks with a No.2 black lead pencil. Enter in the designated blocks your social security number, course code number, edition of course book, lesson or test number, and OPFAC number. Next carefully fill in the numbers opposite the blocks you have just marked. Answer spaces are arranged in vertical sequence. Make only one mark to answer one question. To avoid erasures, at first mark your selection lightly or with a small dot. Then, after you are satisfied no changes will be made, darken circles carefully.		EXAMPLE: WRONG 1 (A) (B) (C) (D) WRONG 2 (A) (B) (C) (D) WRONG 3 (A) (B) (C) (D) RIGHT 4 (A) (B) (C) (D)		Name: <u>BARISTA</u> <u>1</u> <u>M</u> <small>(last) (first) (middle)</small> Rank/Rate: <u>E-9</u> <u>BACM</u> Course title: <u>BA1</u> <u>SERIES 51</u> Present unit address: <u>USCG TRACEN ANYTOWN</u> Zip Code: _____	
PRIVACY ACT OF 1974 A. AUTHORITY FOR USE AND WHETHER MANDATORY OR VOLUNTARY - 14 U.S.C. Sec 633. Voluntary. B. PRINCIPAL PURPOSE - Identify student record. Score test answers. C. ROUTINE USES - Establish score on test. Enter score on student record. D. EFFECT ON YOU IF YOU DO NOT PROVIDE INFORMATION REQUESTED - Your answer sheet will not be scored. You will not achieve course completion.					
SOCIAL SECURITY NUMBER 0 0 0 0 0 0 0 5 (1-9) (1-9) (1-9) (1-9) (1-9) (1-9) (1-9) (1-9)		OPFAC NUMBER (0-9) (0-9) (0-9) (0-9) (0-9) (0-9) (0-9) (0-9)		COURSE CODE 0 9 9 9 (0-9) (0-9) (0-9) (0-9)	
DATE TEST ADMINISTERED MONTH DAY YEAR Jan (1) () () Feb (2) () () Mar (3) () () Apr (4) () () May (5) () () June (6) () () July (7) () () Aug (8) () () Sept (9) () () Oct (10) () () Nov (11) () () Dec (12) () ()		TEST NO. 1 5 1 (0-9) (0-9) (0-9)		DO NOT WRITE IN THIS BLOCK, FOR USE OF U.S. COAST GUARD INSTITUTE (0-9) (0-9) (0-9) (0-9) (0-9) (0-9) (0-9) (0-9)	
NOTE THE RESPONSES ARE IN VERTICAL SEQUENCE					
1 (A) (B) (C) (D)		16 (A) (B) (C) (D)		31 (A) (B) (C) (D)	
2 (A) (B) (C) (D)		17 (A) (B) (C) (D)		32 (A) (B) (C) (D)	
3 (A) (B) (C) (D)		18 (A) (B) (C) (D)		33 (A) (B) (C) (D)	
4 (A) (B) (C) (D)		19 (A) (B) (C) (D)		34 (A) (B) (C) (D)	
5 (A) (B) (C) (D)		20 (A) (B) (C) (D)		35 (A) (B) (C) (D)	
6 (A) (B) (C) (D)		21 (A) (B) (C) (D)		36 (A) (B) (C) (D)	
7 (A) (B) (C) (D)		22 (A) (B) (C) (D)		37 (A) (B) (C) (D)	
8 (A) (B) (C) (D)		23 (A) (B) (C) (D)		38 (A) (B) (C) (D)	
9 (A) (B) (C) (D)		24 (A) (B) (C) (D)		39 (A) (B) (C) (D)	
10 (A) (B) (C) (D)		25 (A) (B) (C) (D)		40 (A) (B) (C) (D)	
11 (A) (B) (C) (D)		26 (A) (B) (C) (D)		41 (A) (B) (C) (D)	
12 (A) (B) (C) (D)		27 (A) (B) (C) (D)		42 (A) (B) (C) (D)	
13 (A) (B) (C) (D)		28 (A) (B) (C) (D)		43 (A) (B) (C) (D)	
14 (A) (B) (C) (D)		29 (A) (B) (C) (D)		44 (A) (B) (C) (D)	
15 (A) (B) (C) (D)		30 (A) (B) (C) (D)		45 (A) (B) (C) (D)	
		46 (A) (B) (C) (D)		61 (A) (B) (C) (D)	
		47 (A) (B) (C) (D)		62 (A) (B) (C) (D)	
		48 (A) (B) (C) (D)		63 (A) (B) (C) (D)	
		49 (A) (B) (C) (D)		64 (A) (B) (C) (D)	
		50 (A) (B) (C) (D)		65 (A) (B) (C) (D)	
		51 (A) (B) (C) (D)		66 (A) (B) (C) (D)	
		52 (A) (B) (C) (D)		67 (A) (B) (C) (D)	
		53 (A) (B) (C) (D)		68 (A) (B) (C) (D)	
		54 (A) (B) (C) (D)		69 (A) (B) (C) (D)	
		55 (A) (B) (C) (D)		70 (A) (B) (C) (D)	
		56 (A) (B) (C) (D)		71 (A) (B) (C) (D)	
		57 (A) (B) (C) (D)		72 (A) (B) (C) (D)	
		58 (A) (B) (C) (D)		73 (A) (B) (C) (D)	
		59 (A) (B) (C) (D)		74 (A) (B) (C) (D)	
		60 (A) (B) (C) (D)		75 (A) (B) (C) (D)	
				76 (A) (B) (C) (D)	
				77 (A) (B) (C) (D)	
				78 (A) (B) (C) (D)	
				79 (A) (B) (C) (D)	
				80 (A) (B) (C) (D)	
				81 (A) (B) (C) (D)	
				82 (A) (B) (C) (D)	
				83 (A) (B) (C) (D)	
				84 (A) (B) (C) (D)	
				85 (A) (B) (C) (D)	
				86 (A) (B) (C) (D)	
				87 (A) (B) (C) (D)	
				88 (A) (B) (C) (D)	
				89 (A) (B) (C) (D)	
				90 (A) (B) (C) (D)	
				91 (A) (B) (C) (D)	
				92 (A) (B) (C) (D)	
				93 (A) (B) (C) (D)	
				94 (A) (B) (C) (D)	
				95 (A) (B) (C) (D)	
				96 (A) (B) (C) (D)	
				97 (A) (B) (C) (D)	
				98 (A) (B) (C) (D)	
				99 (A) (B) (C) (D)	
				100 (A) (B) (C) (D)	
				101 (A) (B) (C) (D)	
				102 (A) (B) (C) (D)	
				103 (A) (B) (C) (D)	
				104 (A) (B) (C) (D)	
				105 (A) (B) (C) (D)	
				106 (A) (B) (C) (D)	
				107 (A) (B) (C) (D)	
				108 (A) (B) (C) (D)	
				109 (A) (B) (C) (D)	
				110 (A) (B) (C) (D)	
				111 (A) (B) (C) (D)	
				112 (A) (B) (C) (D)	
				113 (A) (B) (C) (D)	
				114 (A) (B) (C) (D)	
				115 (A) (B) (C) (D)	
				116 (A) (B) (C) (D)	
				117 (A) (B) (C) (D)	
				118 (A) (B) (C) (D)	
				119 (A) (B) (C) (D)	
				120 (A) (B) (C) (D)	

Note: Image not to scale.

Rating Advancement Test (RAT) (continued)

Sample Printed RAT

The below illustration depicts the first page of a paper, unclassified Rating Advancement Test. If the test is classified be sure to change the header and footer to reflect the classification.

FOR OFFICIAL USE ONLY
BA1 Series 51

BA1 (1st) Series 51

ADMINISTRATION

1. QID 6.2.1/1

Espresso is best brewed _____.

A. between 198°F and 204°F
B. from 30 to 60 seconds
C. from 7 to 10 seconds
D. between 250°F and 212°F

2. QID 6.2.1/2

You are assigned to perform a safety and health inspection for the Cyber Cafe. The Barista prepares a cup of espresso and serves it to a customer. You, as the inspector, observed the brewing temperature to be 188°F and the brew time to be 20 seconds for a single shot. You failed the Barista for preparation of espresso.

Select the reason that you failed the Barista for this process.

A. Brewing time too long
B. Brewing temperature too low
C. Brewing time too short
D. Brewing temperature too high

3. QID 6.1.1/1

Which of the following formulas is used to determine electrical current?

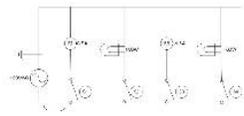
A. $I = P \times E$
B. $E = P \div I$
C. $I = P \div E$
D. $P = I \times E$

4. QID 6.1.1/4

Refer to the schematic below.

S1, S2 and S4 are normally closed producing a continuous load. Periodically, S3 is closed to provide ventilation when the ambient temperature exceeds 125°F.

In order to be in compliance with the 1996 National Electrical Code (NEC) Sec. 384-16(c) concerning sizing of Over Current Protection Devices (OCPD), the minimum acceptable size of the circuit breaker in the schematic below is _____ Amps.



A. 50
B. 30
C. 20
D. 15

5. QID 6.4.3/2

Most commercial jetliners have a service (or certificated) ceiling of about _____.

A. 5,100 meters
B. 42,000 feet
C. 4,200 meters
D. 51,000 feet

6. QID 6.1.1/2

If you have to use a light during nighttime operations, which color light would have the least detrimental effect on night vision?

A. White
B. Blue
C. Yellow
D. Red

1

**GO ON TO NEXT PAGE
FOR OFFICIAL USE ONLY**

Note: Image not to scale.

Rating Advancement Test (RAT) (continued)

Statistical Analysis

A statistical analysis of test and item performance is available on demand for all RATs that are administered online. The local e-testing manager is able to generate various reports regarding test performance for review by the RKM/ARKM and other authorized personnel. For paper tests administered by CGI, the RKM/ARKM should request a statistical analysis from CGI using CGI form 1960 after a new series of the RAT has been in use for at least 6 months to be assured of sufficient sampling. The statistical analysis will assist in identifying the following:

- Miskeyed items
- Items with two or more correct answers or answers that are partially correct
- Topics that participants are performing poorly in
- Items that do not discriminate well
- Items that may be too easy or difficult
- Possible clueing in the stem or key

Crediting Items

Any item that is credited means the participant will receive full credit for the item regardless of which response was chosen. When an item is credited, CGI will NOT make any changes to the paper test booklet and future the test takers will not know the items has been credited. This will most likely result in additional challenges being submitted each time the paper RAT is administered. Credited items will not be graded and all participants receive credit for the item. If more than 10% of the items on the paper test have been credited a new series, NOT edition, of the test must be created.

To credit an item on a paper test, complete a computer key change on the EOCT/RAT Score Key Change Profile Sheet (CGI-2801). Once the action has been completed by CGI, an endorsed copy will be returned to the training source. File the copy along with the master copy of the RAT.

In the case of the online RAT, all items that are credited for other than an incorrectly keyed correct response shall be retired in the database and will not appear on future versions of any test. If any wording has been changed in the item it must be retired so the original version remains in the database for defensibility. The retired item(s) may be deleted I.A.W. chapter 4 of this SOP.

Rating Advancement Test (RAT) (continued)

100% Check by RKM/ARKM

After the printed training source's copy of the RAT and Verified Reference CD are returned to the training source by the CG Institute, the RKM/ARKM must take the RAT and receive a score of 100 percent. The RKM/ARKM must take the "100 percent test" using ONLY the CG Institute's printed copy of the RAT and the corrected Reference CD; using previously prepared answer keys or other keyed copies of the RAT is not permitted. Meticulous attention to this crucial step will guarantee that a completely accurate answer key is used to score the candidate's RAT. The RKM/ARKM shall complete the 100 percent test process on all three paper versions of the current series RAT.

Note: Enter "100 percent test" in the name block and leave the social security number blank. The new RAT series will not be placed into service until each "100 percent test" has been graded by the CG Institute.

Challenges

Challenges from participants are usually received by CGI with received by CGI when the RAT is administered. The participant's challenge will be forwarded to the training source via TACCTS or e-mail. The RKM/ARKM shall respond directly to the participant by e-mail following statement:

“Your challenge to the **XXX** RAT has been received and is under review. If the challenge is valid the item(s) will be credited and your score will be adjusted. Thank you for your interest in ensuring the RAT is valid.”

Under no circumstances shall the test item or responses be discussed with the participant.

Setting RAT Cut (Pass) Score

All RAT cut (pass) scores shall be established by using the Modified Angoff method described in Appendix C of this SOP.

If in the judgment and experience of the test developer the modified Angoff method is not justified, then the TRACEN Commanding Officer may make a written request to FORCECOM (FC-T) to waive any or all of these test development procedures. Only with a written waiver from FORCECOM (FC-T) may a TRACEN deviate from these standards.

Upon collection of an adequate sample of actual test results, the modified Angoff method cut-off scores may be revised to reflect actual responses and difficulties for each test item. In this process, cut-off scores are calculated using the actual data through the modified Angoff method algorithms.

Servicewide Examination (SWE)

Introduction

The Performance, Training and Education Manual, COMDTINST M1500.10 (series), serves as the basis for the development of the servicewide examination (SWE). The manual prescribes the minimum occupational and military standards, expressed in enlisted performance qualifications, for advancement in rate. The SWE is the final phase of the advancement competition. The main goal of the SWE is to identify those candidates best qualified for advancement on the basis of rate-related/required knowledge and performance as well as professional military knowledge and to rank order those already considered minimally qualified. The SWE is a norm-referenced examination. The candidate's name appears on the advancement list in order of the final multiple score, which is a combination of time-in-service, time-in-grade, sea duty points, surf duty points, medals and awards, marks, and SWE score. The SWE is developed by the Rating Knowledge Manager (RKM) for each enlisted specialty and administered under the auspices of the Pay & Personnel Center, Advancements Section (PPC (adv)). A common misconception exists that the Coast Guard Institute (CGI) manages the SWE – they do not.

Components

Each SWE is divided into the following two basic parts:

Part I - Professional. Knowledge and performance relevant to the total field of the rate. Many different section topics particular to the rate may be addressed.

Part II – Enlisted Professional Military Education (E-PME) Requirements common to all ratings in each paygrade.

SWE Construction

Two parallel series of the SWE for each paygrade shall be prepared following the procedures outlines below.

- The test items for a SWE shall be written based on the RPQ tasks. There shall be sufficient test items (a minimum of five per task) for the test items database to randomly select enough items to create two separate tests for each paygrade.
 - Select items that have been adequately reviewed. The items selected from your test item database for the SWE may test the same concepts as the RAT. The test items for the SWE shall be written differently than what has already been used.
-

Servicewide Examination (SWE) (continued)

SWE Construction (continued)

The following requirements are specified for writing SWE items:

- All SWEs will be created by appropriate rating SMEs or SMSs.
- Under no circumstances will individuals have access to any examination component (test items) for examinations in which they could participate.
- Each SWE consists of 150 items (questions).
- All items must be written in the four-response, multiple-choice format.
- Test item writing is a precise type of writing. In the ideal workflow, SWE test items are reviewed by two individuals, one with the role of Writer/Editor (W/E) and one with the role of Instructional Systems Specialist (ISS). These roles can be filled by one person if necessary.
- Every SWE item in the Professional and E-PME parts must test a performance-based qualification required by COMDTINST M1414.8 (series) of the target rate or any lower rate.
- Two parallel but different exams for each pay grade (E-5, E-6, etc.) must be developed for each exam cycle. Section emphasis must be identical for each exam for that pay grade.
- Series numbers for each exam cycle are provided by PPC (adv) and should be labeled as indicated in table 6-5 unless otherwise directed by PPC (adv).
- Every item must have a written reference, which confirms the correctness of the item. Each reference must be either:
 - An official directive applicable throughout the Coast Guard
 - **Examples:**
 - Coast Guard publications Commandant Instruction or other documents listed in COMDTNOTE 5600
 - A generally accepted source of information such as *Dutton's Navigations and Piloting*, *Gray's Anatomy*, or a manufacturer's instruction book which would be available to all **and is a required RPQ reference**

Servicewide Examination (SWE) (continued)

Predetermined Test Length for SWE

The Servicewide Examination (SWE) has an established test length of 150 items and maximum allotted administration time of three hours.

The SWE contains a section that tests professional, rate specific test items based upon the Rating Performance Qualifications (RPQs) as well as a section that is based upon Enlisted Professional Military Education Requirements (EPME). Table 6-4 establishes the required section sizes and responsible test item author/exam writer.

Section	Pay Grade	Number of test items	Responsibility
Rating Performance Qualification (RPQ)	E-5	120	Written by appropriate rating RKM. Reviewed by writer/editor (W/E) (or training specialist) and Instructional System Specialist (ISS). (These roles can be filled by one person if necessary.)
	E-6/E-7	125	
	E-8/E-9	80	
Enlisted Professional Military Education Requirement (EPME)	E-5	30	Written by E-PME section, CGTRACEN Petaluma. Reviewed by writer/editor (W/E) (or training specialist) and Instructional System Specialist (ISS). (These roles can be filled by one person if necessary.)
	E-6/E-7	25	
	E-8/E-9	70	

Table 6-4

Section Requirements

Items in each part of the examination should be arranged in specific topic areas called sections. Sections within each part must meet the following requirements:

- Must be homogeneous (i.e., All items in a sections must refer to a common topic corresponding to the RPQ section titles.)
- Items must be referenced to a current RPQ
- Items must have a valid reference
- Sections should be continuous without a page break between them to ensure there is no extra whitespace on each page.
- Ensure the columns do not break mid test item. (e.g. stem and alternatives A & B in left column and C &D in right column)

Servicewide Examination (SWE) (continued)

Procedure

The SWE is developed to sample the Enlisted Rating Performance Qualifications (RPQs). The procedure is listed below:

Step	Action	Responsibility
1	Validate database test items, RPQs, and references	Appropriate RKM or W/E or ISS
2	Prepare exam strategy and develop test section emphasis for next SWE cycle	Appropriate RKM in consultation with Rating Force Master Chief (RFMC)
3	Print/format exams	Appropriate RKM with assistance of testing system administrator
4	Spell-check and proofread the exams	Appropriate RKM and W/E
5	Conduct first edit Note: The ISS and W/E must approve all test items before they are used on an SWE.	W/E and/or ISS
6	Review and make appropriate changes	Appropriate RKM
7	Print camera-ready copy (CRC)	Appropriate RKM with assistance of testing system administrator
8	Conduct camera-ready review	W/E or ISS
9	If revision needed, go to step 6	Appropriate RKM

Servicewide Examination (SWE) (continued)

Procedure (continued)

Step	Action	Responsibility
10	Complete section title sheets (PPC-1400)	Appropriate RKM
11	Prepare answer sheets	Appropriate RKM
12	Dual proofread answer keys with another RKM or W/E	Appropriate RKM
13	Conduct final review	ISS
14	Send servicewide exams to (PPC) adv	Appropriate RKM or local Test Development Manager

Note: Depending on training source and staffing billets, responsibilities of the ISS and/or W/E could be performed by other staff members.

SWE Schedule

The schedules for administering the SWE are outlined in the table below.

SWE Schedule by Paygrade and Service		
Active Duty	E-5 through E-6	November
	E-5 through E-9	May
Reserve Forces	E-5 through E-9	October

Servicewide Examination (SWE) (continued)

Deadlines for SWE Development Dates

To ensure PPC (adv) has sufficient time to review and build the examination packages for administration of the SWE, the dates for delivery of tests and the series numbers through May 2026 are listed in the below table.

SWE Date	Series Number		Type	Due for Local ISS Review	Due in PPC (adv)	Paygrades Tested
	Primary	Substitute				
OCT 2015	90	91	Reserve	01MAY2015	15MAY2014	E5 – E9
NOV 2015	92	93	Active	01JUN2015	15JUN2015	E5 – E6
MAY 2016	94	95	Active	01DEC2015	15DEC2015	E5 – E9
OCT 2016	96	97	Reserve	01MAY2016	15MAY2016	E5 – E9
NOV 2016	98	99	Active	01JUN2016	15JUN2016	E5 – E6
MAY 2017	01	02	Active	01DEC2016	15DEC2016	E5 – E9
OCT 2017	03	04	Reserve	01MAY2017	15MAY2017	E5 – E9
NOV 2017	05	06	Active	01JUN2017	15JUN2017	E5 – E6
MAY 2018	07	08	Active	01DEC2017	15DEC2017	E5 – E9
OCT 2018	09	10	Reserve	01MAY2018	15MAY2018	E5 – E9
NOV 2018	11	12	Active	01JUN2018	15JUN2018	E5 – E6
MAY 2019	13	14	Active	01DEC2018	15DEC2018	E5 – E9
OCT 2019	15	16	Reserve	01MAY2019	15MAY2019	E5 – E9
NOV 2019	17	18	Active	01JUN2019	15JUN2019	E5 – E6
MAY 2020	19	20	Active	01DEC2019	15DEC2019	E5 – E9
OCT 2020	21	22	Reserve	01MAY2020	15MAY2020	E5 – E9
NOV 2020	23	24	Active	01JUN2020	15JUN2020	E5 – E6
MAY 2021	25	26	Active	01DEC2020	15DEC2020	E5 – E9
OCT 2021	27	28	Reserve	01MAY2021	15MAY2021	E5 – E9
NOV 2021	29	30	Active	01JUN2021	15JUN2021	E5 – E6
MAY 2022	31	32	Active	01DEC2021	15DEC2021	E5 – E9
OCT 2022	33	34	Reserve	01MAY2022	15MAY2022	E5 – E9
NOV 2022	35	36	Active	01JUN2022	15JUN2022	E5 – E6
MAY 2023	37	38	Active	01DEC2022	15DEC2022	E5 – E9
OCT 2023	39	40	Reserve	01MAY2023	15MAY2023	E5 – E9
NOV 2023	41	42	Active	01JUN2023	15JUN2023	E5 – E6
MAY 2024	43	44	Active	01DEC2023	15DEC2023	E5 – E9
OCT 2024	45	46	Reserve	01MAY2024	15MAY2024	E5 – E9
NOV 2024	47	48	Active	01JUN2024	15JUN2024	E5 – E6
MAY 2025	49	50	Active	01DEC2024	15DEC2024	E5 – E9
OCT 2025	51	52	Reserve	01MAY2025	15MAY2025	E5 – E9
NOV 2025	53	54	Active	01JUN2025	15JUN2025	E5 – E6
MAY 2026	55	56	Active	01DEC2025	15DEC2025	E5 – E9

Table 6-5

Servicewide Examination (SWE) (continued)

SWE Package

After the test development process is completed, mail the following completed items by FedEx to PPC (adv). Sample items required in the SWE package are included on the following pages.

- Properly formatted, camera-ready copy of the SWE. Format will depend on whether the test is unclassified or classified.
- The professional section of the SWE must end on an even-numbered page; therefore, the test may require a blank page.
- SWE Section Title Sheet (PPC-1400)
- SWE Answer Sheet (PPC -4801) - the examination answer key is prepared on this form.

Packaging and Mailing Instructions

All testing material, whether sensitive, FOUO or classified, shall be mailed/shipped double wrapped with the sealed, inside envelope containing the testing material marked clearly on both sides with the following:

SENSITIVE MATERIALS TO BE OPENED BY A TESTING MATERIAL OFFICER ONLY
--

The letters shall be a minimum of ¼ inch in height. The office and person designated to receive the testing material shall be identified on the inner envelope. There shall be no indication on the outer packaging as to the contents of the inner packaging.

When sensitive testing material is mailed between the training source (TS) and PPC (adv), it must be accounted for by signature. Unless otherwise directed the services of Federal Express (FedEx) shall be used for all mailings of testing material. Do NOT use certified mail. If mailing classified material follow current instructions in COMDTINST M5510.23A (Classified Information Management Program).

Servicewide Examination (SWE) (continued)

SWE Section Illustrations

The following sections illustrate properly formatted components of the SWE package. Note the default print font size of the Word template used to print the SWE is 10 pt Arial. The template designed for the SWE shall be utilized when printing the SWE and is available on the CG Portal SharePoint site.

Unclassified SWE page

Sample unclassified SWE page

FOR OFFICIAL USE ONLY
U.S. COAST GUARD SERVICEWIDE EXAM

BA1 Series 84

ADMINISTRATION

1. QID 4.2.1/1

Espresso is best brewed ____.

A. from 30 to 60 seconds
B. from 7 to 10 seconds
C. between 250°F and 212°F
D. between 198°F and 204°F

2. QID 6.2.1/2

You are assigned to perform a safety and health inspection for the Cyber Cafe. The Barista prepares a cup of espresso and serves it to a customer. You, as the inspector, observed the brewing temperature to be 188°F and the brew time to be 20 seconds for a single shot. You failed the Barista for preparation of espresso.

Select the reason that you failed the Barista for this process.

A. Brewing temperature too low
B. Brewing time too short
C. Brewing temperature too high
D. Brewing time too long

3. QID 4.1.1/1

Which of the following formulas is used to determine electrical current?

A. $I = P \times E$
B. $I = P \div E$
C. $P = I \times E$
D. $E = P \div I$

4. QID 6.1.1/4

Refer to the schematic below.

S1, S2 and S4 are normally closed producing a continuous load. Periodically, S3 is closed to provide ventilation when the ambient temperature exceeds 125°F.

In order to be in compliance with the 1996 National Electrical Code (NEC) Sec. 384-16(c) concerning sizing of Over Current Protection Devices (OCPD), the minimum acceptable size of the circuit breaker in the schematic below is ____ Amps.

A. 50
B. 30
C. 15
D. 20

5. QID 5.4.3/2

Most commercial jetliners have a service (or certificated) ceiling of about ____.

A. 42,000 feet
B. 4,200 meters
C. 5,100 meters
D. 51,000 feet

6. QID 4.1.1/2

If you have to use a light during nighttime operations, which color light would have the least detrimental effect on night vision?

A. Blue
B. Red
C. Yellow
D. White

1

GO ON TO NEXT PAGE
FOR OFFICIAL USE ONLY

Note: Image not to scale.

Servicewide Examination (SWE) (continued)

**Classified SWE
Page**

Sample Classified SWE Page.

CONFIDENTIAL
U.S. COAST GUARD SERVICEWIDE EXAM

BA1 Series 84

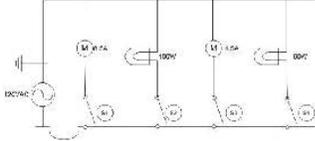
ADMINISTRATION

1. QID 4.2.1/1
Espresso is best brewed _____.
A. from 30 to 60 seconds
B. from 7 to 10 seconds
C. between 250°F and 212°F
D. between 198°F and 204°F

2. QID 6.2.1/2
You are assigned to perform a safety and health inspection for the Cyber Cafe. The Barista prepares a cup of espresso and serves it to a customer. You, as the inspector, observed the brewing temperature to be 188°F and the brew time to be 20 seconds for a single shot. You failed the Barista for preparation of espresso.
Select the reason that you failed the Barista for this process.
A. Brewing temperature too low
B. Brewing time too short
C. Brewing temperature too high
D. Brewing time too long

3. QID 4.1.1/1
Which of the following formulas is used to determine electrical current?
A. $I = P \times E$
B. $I = P \div E$
C. $P = I \times E$
D. $E = P \div I$

4. QID 6.1.1/4
Refer to the schematic below.
S1, S2 and S4 are normally closed producing a continuous load. Periodically, S3 is closed to provide ventilation when the ambient temperature exceeds 125°F.
In order to be in compliance with the 1996 National Electrical Code (NEC) Sec. 384-16(c) concerning sizing of Over Current Protection Devices (OCPD), the minimum acceptable size of the circuit breaker in the schematic below is ____ Amps.

5.

A. 50
B. 30
C. 15
D. 20

6. QID 5.4.3/2
Most commercial jetliners have a service (or certificated) ceiling of about _____.
A. 42,000 feet
B. 4,200 meters
C. 5,100 meters
D. 51,000 feet

7. QID 4.1.1/2
If you have to use a light during nighttime operations, which color light would have the least detrimental effect on night vision?
A. Blue
B. Red
C. Yellow
D. White

1

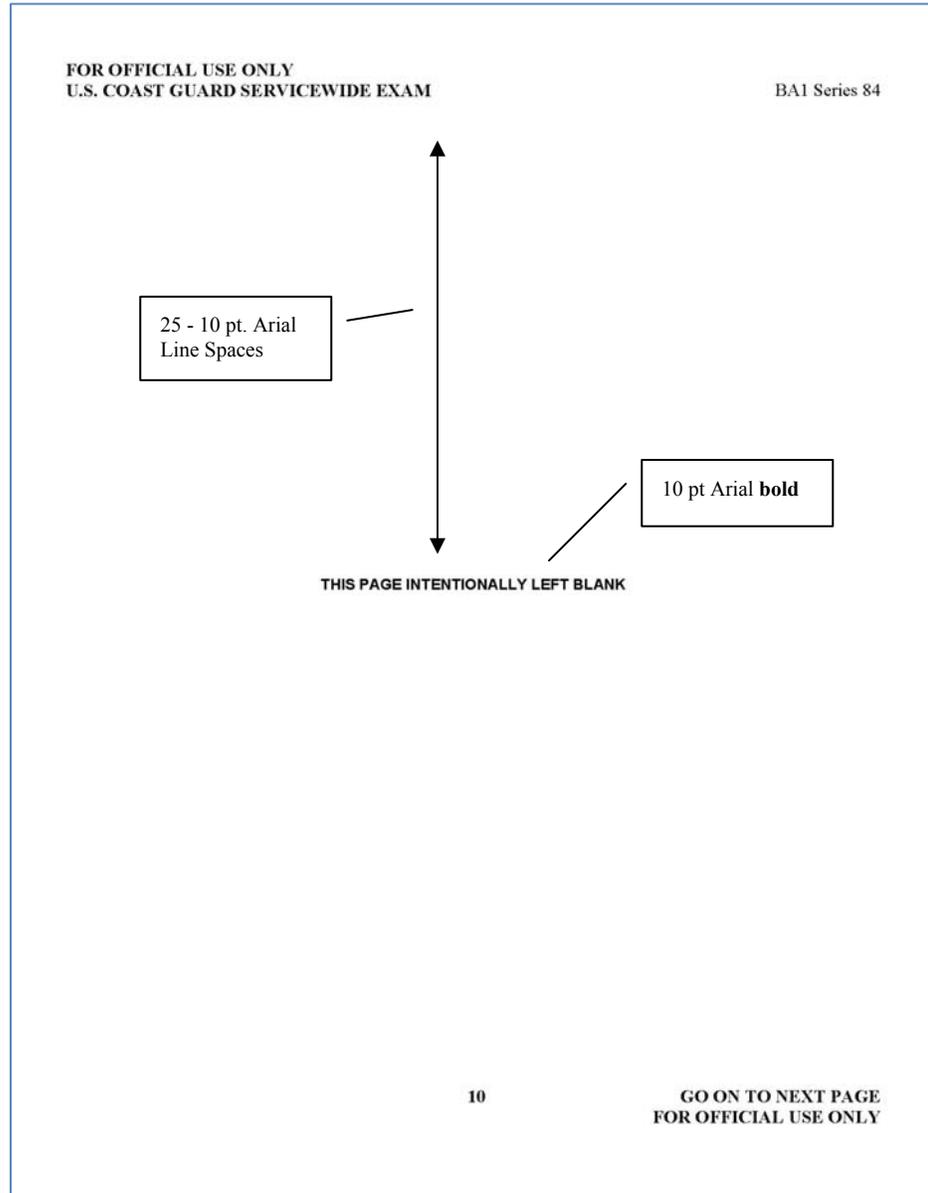
GO ON TO NEXT PAGE
CONFIDENTIAL

Note: Image not to scale.

Servicewide Examination (SWE) (continued)

SWE Blank Page

Sample SWE blank page to be inserted if printout ends on an odd page.



Note: Image not to scale.

Servicewide Examination (SWE) (continued)

SWE Section Title Sheet

Each SWE is divided into sections. Each section title should describe the topic(s) included in the section. The section title sheet (PPC-1400) is used to record section titles, as shown below, and is prepared by the RKM/test writer after completion of the dual-column paper examination. The following applies:

- Limit each section title to a maximum of 30 characters including spaces between words
- Maximum section titles, including the E-PME section, are limited to no more than 12. If the RPQs contain more than 12 units that are being tested you must combine units in the section titles.
 - **Example:** ADMINISTRATION AND SAFETY
- Include “E-PME” line with the appropriate number of test items for the paygrade filled in
- Ensure the TOTAL equals 150
- Ensure the “Question Numbers” do not overlap
 - Example: 1 – 10, 11 – 15, etc. NOT 1 – 10, 10 – 15, etc.

SWE SECTION TITLE WORKSHEET																																			
RATE:		BACM																															FOR ADV USE ONLY		
SERIES:		84																																	
PAY GRADE:		E-9																																	
///	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	QUESTIONS PER SECTION	QUESTION NUMBERS <small>(Ex. 1-30, 31-50)</small>	DA SECTION NUMBER		
01	A	D	M	I	N	I	S	T	R	A	T	I	O	N																			30	1-30	
02	S	A	F	E	T	Y																										15	31-45		
03	M	A	I	N	T	E	N	A	N	C	E																					20	46-65		
04	T	R	A	I	N	I	N	G																								15	66-80		
05	E	P	M	E																												70	81-150		
06																																			
07																																			
08																																			
09																																			
10																																			
11																																			
12																																			
TOTAL:																												150							
* PRINT TITLES ONE LETTER PER BLOCK AND LEAVE A SPACE BETWEEN WORDS. * ENSURE EPME TITLE IS ADDED AS THE LAST SECTION SO QUESTIONS TOTAL 150 * PROFILE LETTERS CANNOT ACCEPT MORE THAN 30 CHARACTERS FOR A SECTION TITLE. * ABBREVIATE, WHEN NECESSARY, TO ENSURE ONLY 30 CHARACTERS PER SECTION TITLE.																																			
DHS-US Coast Guard PPC FORM 1400 Revised 12/7/2011																				BACM I. M. BARISTA EXAM WRITER										01/07/2014					

Servicewide Examination (SWE) (continued)

SWE Rate Codes

The examination rate code is entered on the exam answer key as the first three numbers in the “EXAM IDENTIFICATION NO.” block. The SWE rate codes are listed in table 6-6 below.

SWE RATING SPECIALTY CODES							
Rating	Code	Rating	Code	Rating	Code	Rating	Code
AMT2	206	EM2	219	IT2	280	PA2	236
AMT1	106	EM1	119	IT1	180	PA1	136
AMTC	006	EMC	019	ITC	080	PAC	036
AMTCS	806	EMCS	819	ITCS	880	PACS	836
AMTCM	906	EMCM	919	ITCM	980	PACM	936
AET1	202	ET2	222	IV2	226	SK2	250
AET2	102	ET1	122	IV1	126	SK1	150
AETC	002	ETC	022	IVC	026	SKC	050
AETCS	802	ETCS	822	IVCS	826	SKCS	850
AETCM	902	ETCM	922	IVCM	926	SKCM	950
AST2	210	FS2	251	ME2	240	YN2	275
AST1	110	FS1	151	ME1	140	YN1	175
ASTC	010	FSC	051	MEC	040	YNC	075
ASTCS	810	FSCS	851	MECS	840	YNCS	875
ASTCM	910	FSCM	951	MECM	940	YNCM	975
BM2	212	GM2	229	MK2	232		
BM1	112	GM1	129	MK1	132		
BMC	012	GMC	029	MKC	032		
BMCS	812	GMCS	829	MKCS	832		
BMCM	912	GMCM	929	MKCM	932		
DC2	215	HS2	230	MST2	234		
DC1	115	HS1	130	MST1	134		
DCC	015	HSC	030	MSTC	034		
DCCS	815	HSCS	830	MSTCS	834		
DCCM	915	HSCM	930	MSTCM	934		
DV2	282	IS2	290	OS2	238		
DV1	182	IS1	190	OS1	138		
DVC	082	ICS	090	OSC	038		
DVCS	882	ISCS	890	OSCS	838		
DVCM	982	ISCM	990	OSCM	938		
RATING DEFINITIONS							
AMT	Aviation Maintenance Technician	IS	Intelligence Specialist	IT	Information Systems Technician	IV	Investigator
AET	Avionics Electrical Technician	ME	Maritime Enforcement Specialist	MK	Machinery Technician	MST	Marine Science Technician
AST	Aviation Survival Technician	OS	Operations Specialist	PA	Public Affairs Specialist	SK	Storekeeper
BM	Boatswain's Mate	PA	Public Affairs Specialist	SK	Storekeeper	YN	Yeoman
DC	Damage Controlman	SK	Storekeeper				
DV	Diver	YN	Yeoman				
EM	Electrician's Mate						
ET	Electronics Technician						
FS	Food Service Specialist						
GM	Gunner's Mate						
HS	Health Services Technician						

Table 6-6

Servicewide Examination (SWE) (continued)

SWE Answer Key Header

The RKM must prepare an answer key for the professional (rating) portion of each examination utilizing the USCG Examination Answer Sheet (PPC-4801) revised 04/2010. The following steps shall be followed:

Step	Action									
1	Use #2 soft lead black pencil									
2	Fill in the name of the person who is preparing the answer key: Example: Barista, I. M.									
3	Fill in the paygrade of the exam in the Rate/Grade space. Example: E-5									
4	Fill in the exam short title. Example: AET2									
5	Leave the Social Security Number blank									
6	<p>In the Exam Identification No. block enter the following:</p> <ul style="list-style-type: none"> • Three number rate code (see table 6-5) • Exam Series number (see table 6-4) • The “0” is already coded • 999 as the last three numbers <p>Example: the AET2 series 84 exam would be coded as:</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td>2</td><td>0</td><td>2</td><td>8</td><td>4</td><td>0</td><td>9</td><td>9</td><td>9</td> </tr> </table>	2	0	2	8	4	0	9	9	9
2	0	2	8	4	0	9	9	9		
7	Blacken the corresponding spaces below the Exam Identification Block									
8	Blacken the appropriate answer spaces next to each test item number									

Servicewide Examination (SWE) (continued)

Prescoring Audit

The RKM must audit the entire examination as soon as possible after administration. Every test item (stem and all alternatives) must be verified to ensure they are valid as of the day of the examination to account for any changes in the references used to generate the items. Any challenges received from candidates must be considered during the pre-scoring audit (dirty run). The audit consists of two phases. Phase one is accomplished before receipt of the dirty run and challenges. Phase two is accomplished after receipt of the dirty run.

- The following shall be accomplished during Phase one:

Step	Action
1	Review all items based on references that have been changed between the date the examination was prepared and the date it was administered
2	Complete as much item verification and research as possible prior to receipt of the “dirty run” due to the five working day turnaround required
3	Complete the <i>Credit and/or Score Key Change Form</i> for any item that was not valid on the day of the examination

- The following shall be accomplished during Phase two:

Step	Action										
1	Review every item that has a “P-value” higher than the “P-value” for the correct response or has a value of 1.000 or 0.000.										
2	Review every item that has a negative “d-Value” for the correct response or a positive “d-value” for an incorrect response										
3	Check scoring key and take the following action: <table border="1" data-bbox="690 1564 1404 1707"> <thead> <tr> <th>IF...</th> <th>THEN...</th> </tr> </thead> <tbody> <tr> <td>score key was miskeyed</td> <td>correct the score</td> </tr> <tr> <td>score key is correct</td> <td>check item wording for accuracy</td> </tr> <tr> <td>item is unclear or irrelevant</td> <td>credit the item</td> </tr> <tr> <td>item is clear and relevant</td> <td>take no action</td> </tr> </tbody> </table>	IF...	THEN...	score key was miskeyed	correct the score	score key is correct	check item wording for accuracy	item is unclear or irrelevant	credit the item	item is clear and relevant	take no action
IF...	THEN...										
score key was miskeyed	correct the score										
score key is correct	check item wording for accuracy										
item is unclear or irrelevant	credit the item										
item is clear and relevant	take no action										

Servicewide Examination (SWE) (continued)

SWE Audit Results

The RKM shall utilize the *Credit and/or Score Key Change Form* to record the SWE audit results. The completed form shall be submitted to PPC (adv) with the appropriate annotations to indicate items that will or will not be credited and/or any changes to the answer key.

SWE Credit and/or Score Key Change Form

Upon final review and audit the RKM must complete the *Credit and/or Score Key Change Form* and forward it to PPC (adv). Even if credit to a challenge is not given, the RKM must complete and forward the form. The completed form may be e-mailed via encrypted e-mail to appropriate personnel designated by PPC (adv). The illustration below depicts the first page of the *Credit and/or Score Key Change Form*. Page two is a continuation of the table and a signature block.

U.S. Department of
Homeland Security
United States
Coast Guard

1418

MEMORANDUM

From: BACM I.M. BARISTA Reply to: I. M. BARISTA
Printed Name of Subject Matter Specialist Attn of: 555-555-1212
POC NAME/PHONE

To: D. C. ROSE
CG PPC (ADV)

Subj: CREDITS AND CHANGES TO ANSWER KEYS FOR THE OCT 2014 SWE
Month/Yr

1. There will be **NO** credits or answer key changes for the OCT 2014 SWE for the following series and exam rates:

Series: 84 Exam Rates: BAC BACS BACM

Series: 85 Exam Rates: BA2 BA1 BAC BACS BACM

2. Please change the answer key for the OCT 2014 SWE as follows:
Month/Yr.

CHANGES TO SWE ANSWER KEY					
<small>(Chose either to throw a question out OR to change the correct answer choice but not both)</small>					
Series	Rating	Question #	Throw out the question and give all candidates credit	Change current answer FROM (A, B, C, or D)	TO (A, B, C, or D)
84	BA2	5	CREDIT		
84	BA1	68	CREDIT		
84	BA1	69	CHANGE	A	B

Servicewide Examination (SWE) (continued)

SWE Credit and/or Score Key Change Form (continued)

Although most correspondence challenging the validity of the content of the SWE will be received by PPC (adv), some challenges may be sent directly to the RKM. The RKM shall review any direct challenges during the pre-scoring audit (dirty run) stage. In the rare instance of the RKM receiving a direct challenge, PPC (adv) shall be notified immediately using the *Credit and/or Score Key Change Form*. Should the direct challenge be received after the dirty run closeout, the RKM or the local Test Development Manager shall contact PPC (adv) for guidance. Additionally, the RKM shall acknowledge receipt of the direct challenge to the source of the challenge (SWE Officer at the Exam Board Unit) using the Examination Question Inquiry memo shown below. The RKM shall not provide any information concerning the credit decision or discuss the contents of the item(s).

U.S. Department of Transportation United States Coast Guard 	Commanding Officer U.S. Coast Guard NAME OF TRAINING SOURCE	ADDRESS CITY, STATE ZIP Staff Symbol: Phone: # Fax: Email: RKM@uscg.mil
		5600 DATE
MEMORANDUM		
From: <i>RKM name & signature</i> NAME OF TRAINING SOURCE	Reply to: <i>Staff Symbol</i> Attn of: <i>RKM Name</i>	
To: Commanding Officer (<i>Name of Exam Board Unit</i>) Attn: <i>Name of SWE Officer</i>		
Subj: SWE INQUIRY		
Ref: (a) Your SWE Challenge dtd DD MMM YYYY		
<ol style="list-style-type: none"> <li data-bbox="581 1092 1169 1144">1. Your correspondence concerning item(s) XX on the XXX servicewide examination for <i>TITLE OF PERSON SUBMITTING CHALLENGE</i> (Ex. BA2 I Failed) has been received and reviewed against the XX Rating Performance Qualifications (RPQs). <li data-bbox="581 1155 1193 1186">2. Each challenged test item is checked for accuracy and validity. If an examination question is determined to be invalid, that item is credited for all examination participants. <li data-bbox="581 1197 1177 1270">3. Your effort to ensure a fair and accurate examination is appreciated. To prevent compromises, the Rating Knowledge Manager (RKM) does not provide answers or references for the examination items. However, if any other information would be helpful, please contact me directly, at the address, e-mail or telephone number above. <li data-bbox="581 1281 1128 1333">4. You are encouraged to submit future challenges directly to PPC (adv) following the challenge procedures provided to the SWE Officer and the candidates at the time of administration of the SWE. 		
Copy: PPC (adv)		

Servicewide Examination (SWE) (continued)

Preparation of SWE in the Testing Software

The SWE test items and examination are prepared in much the same manner as the RAT within the USCG enterprise testing software. The SWE items should be designed to test general knowledge of the entire rating at and below the paygrade being tested at a higher cognitive level than the RAT. All items on the SWE must be linked to the current RPQs and associated tasks have a valid reference.

Standard SWE Topic Structure

It is important to establish a standard topic structure when developing test items within the CG enterprise testing application database. By maintaining a standard topic structure it is possible to select test items at the lowest level desired as well as perform analytics by specific topic. Figure 6-4 depicts the standard SWE structure.

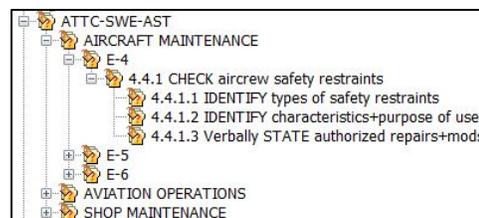


Figure 6-4

Notice that the training command responsible for the rating is always the first part of the “root” topic – in the example; ATTC represents Aviation Technical Training Center. The next part of the root topic name identifies this as a SWE which is followed by the rating short title.

The sub-topic structure shall be set to reflect the current RPQ naming/structure schema as shown in Figure 6-4. If there are more than nine (9) units, RPQs, EOs or Tasks the first number should be preceded by a zero (0) to provide logical sorting, e.g. Task 4.4.1.1 may become Task 4.04.1.1 if unit 4 of the E-4 RPQ 4.1.1 had a task 10.1.1. This would place task 10 after task 09 in the list instead of after task 1.

Access to SWE topic folders is limited to the TRACEN testing system administrator and assistant, the RKM, Assigned Writer/Editor (W/E) and Instructional Systems Specialist (ISS).

Servicewide Examination (SWE) (continued)

Question Identifier (QID)

To ensure positive identification and provide a means for candidates to reference a test item should they want to submit a challenge to PPC, each test item must have a Question Identifier (QID) assigned which shows the task and the number of the item (after the “/”). Figure 6-5 illustrates how the QID appears in the database by description and how it appears to the candidate on the computer screen. (Actual item contents have been hidden). The QID also makes it easier to locate the item in the database.

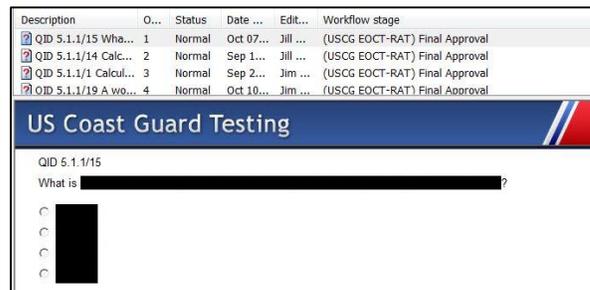


Figure 6-5

Assessment Folder Naming

The SWE shall be developed within the current CG enterprise testing application database and maintained within assessment folders. Separate folders will be established for each rating specialty with access limited to the TRACEN testing system administrator and assistant, RKM, Assigned Writer/Editor (W/E) and Instructional Systems Specialist (ISS). Assessment folders shall maintain a standard naming convention as illustrated by Figure 6-6 showing TRACEN ownership-SWE-Rate short title.

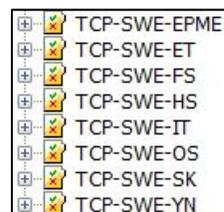


Figure 6-6

Sub folders shall be created for each series of the SWE as illustrated by figure 6-7.

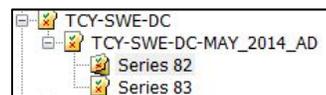


Figure 6-7

Servicewide Examination (SWE) (continued)

SWE Naming

The SWE naming convention within the assessment folder shall be standard throughout the Coast Guard as illustrated by figure 6-8. This is the title that will appear on the paper copy of the SWE.

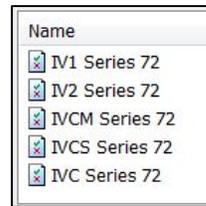


Figure 6-8

SWE Construction

Designing the SWE is accomplished much the same as designing the RAT; the primary difference is that a difficulty level will not be assigned via metatag use. Once all ratings in the USCG are utilizing the enterprise testing application to develop SWEs and PPC (adv) is utilizing it to score and gather statistical data on test items, a statistical difficulty level can be assigned to each item.

Refer to the RAT construction for step-by-step guidance to design the SWE. Modify the steps accordingly to omit applying difficulty metatags.

Printing the SWE

PPC (adv) requires paper versions of the SWE. These paper versions must be printed single sided using the current template for formatting as approved by PPC (adv). Paper versions may be printed locally and mailed to PPC (adv). Future plans are for PPC (adv) to print the SWE on demand directly. Contact your local testing system administrator who will be your liaison between you and PPC (adv) when you are ready to print.

THIS PAGE LEFT INTENTIONALLY BLANK

Chapter 7

PERFORMANCE TESTS

Performance Test Development

This chapter provides amplifying information regarding the development of performance tests as outline in TRASYS SOP Volume 5 (Chapter 5). Until such time that Volume 5 is updated to reference this SOP as the authority for performance test development, Volume 5 shall be followed and augmented with guidance contained in this chapter.

Mandate

In accordance with SOP 3, Performance tests should simulate the standards expressed in the conditions of the TPO, using the highest level of simulation possible.

Knowledge-based tests are appropriate only when specified within the TPOs. For instance, when the performance requires “writing” or “calculating a number” or “recalling from memory without references”, the test of the TPO will likely be a paper-based test.

Performance tests are always preferred over knowledge-based (predictive) tests for demonstration of skill mastery.

Introduction to Performance Test Development

Overview of Performance Testing

Performance tests are the best method for the accurate assessment of an individual's overall competency. In order to develop good performance test items, ensure that the objectives are sufficiently specific, determine the level of test fidelity possible, determine scoring procedures, write the items and instructions, validate items and instructions, obtain approval, and update the Student Evaluation Plan.

Performance tests require the learner to perform an overt action or series of actions, rather than verbalize or write (unless the required performance is speaking or writing).

Written vs. Performance Test Items

The major differences between written tests and performance tests are:

- Written test items
 - Primarily abstract or verbal
 - Items address knowledge and content
 - Items usually address independent aspects
 - Performance test items
 - Primarily non-verbal
 - Items are skills, performance, or job related decisions
 - Items sequentially presented and errors early in the sequence may affect later items
-

Collection of Documentation

Prerequisite to Test Development

In order to begin the performance test development process, collect and have available all necessary documentation, policies, and procedures.

Some examples of this documentation include:

- TPOs
- Task analysis
- Evaluation criteria selection
- Job aids
- Checklists
- Coast Guard procedural guidance
 - COMDTINST, SOP, NAVSEA, etc.
- EPSS

For information about documentation used to develop performance tests, see SOP 5, Chapter 5, Section 5.3.

Review Performance Objectives

Whether developing a written test, or a performance test, a critical task is deciding not what to test, but how to best test the objective.

Usually, the performance objective clearly indicates what performance is demonstrated, how it is judged as correctly performed, the conditions of performance, and number of successful iterations of performance required for mastery. If the performance objective includes all these factors, the task of the performance test developer becomes very clear—duplicate in the test item, with as high fidelity (match) as possible, the performance (action) called for in the objective, under the conditions called for in the objective (or a sampling of conditions), and evaluate the performance, using the criteria in the standard of the objective.

Collection of Documentation (continued)

Reliance on Results of Analysis

The test developer relies on the results of the design phase of training development to create the performance assessment. The results of the formal task analysis and task detailing, for example, as well as the TPOs that are determined by the analysis and the constraints and parameters that are revealed by it, are integral inputs to determining content. A great deal of the design of a good performance test item is directly attributable to the quality of the TPOs. Upon occasion, a TPO may not meet the needs of the performance test developer. It may not be specific enough or, indeed, it might be too specific for integration into a performance test. In these cases, it is important for the developer to coordinate with the Instructional System Specialist (ISS).

Designing Performance Test Items

Steps for Designing a Performance Test

To design performance test items:

Step	Action
1	Identify the performance in each objective.
2	Draft a criterion-referenced test, specifying the performance required
3	Identify the conditions under which the performance should occur (If they cannot be emulated, match them as closely as possible.)
4	Add the standards, and any other evaluation criteria to the test (refer to SOP 5, Design Phase, WS-J.2)
5	Establish the steps for successful evaluation of this task (process and/or product) (refer to Analysis task details, SOP 5, WSF.1)
6	Define the testing criteria and instructions for administering the performance tests

Level of Fidelity

It is often impractical to reproduce a real job situation, or provide actual equipment. However, simulate critical job elements, via various means, for training and testing purposes. For example, an electronics technician is required to inspect circuits and repair malfunctions on a piece of simulated equipment. Some of the reality of the work setting is sacrificed, but the critical job elements—namely, the wiring of the components found in complex electronic equipment from the technical reference—are present on paper; thus the test is readily recognized as a realistic representation of the tasks encountered on the job.

The match between the TPO, the actual test item, and on-the-job performance is called fidelity. The TPO should indicate a high degree of fidelity (match) with the performance required on the job. Ensure the test items display a high degree of fidelity with the TPO.

Designing Performance Test Items (continued)

Factors in Determining Level of Fidelity

To determine the level of fidelity, consider such factors as the time required to perform a given task, the type of equipment required, the ability to present the task in a uniform (standard) manner, risk of damage to equipment or personnel, and the ability to evaluate an individual's performance with a high degree of objectivity.

These considerations impose realistic constraints on the level of fidelity of the performance test item. Often, compromises are in order. Instead of requiring performance of a complex task, decide to limit the test to one or two phases of the task, such as preparing only one slide of a biological specimen, but identifying a larger number of mounted specimens. At times, computer simulations are used as a compromise.

Product Evaluation

In a performance test, the learner actually performs a task and is judged against predetermined criteria. A performance test may involve product evaluation, process evaluation, or both.

Product and/or Process Evaluation

Product evaluation is always appropriate if the objective specifies a product. When a product measure is required, incorporate it into the TPO and carry it over into the test items.

Use product measurement when:

- The objective specifies a product as output
 - The product is measured as either presence or characteristics, such as voltage, length, etc.
 - The procedure leading to the product can vary, without affecting the product
-

Designing Performance Test Items (continued)

Process Measurement

Process measurement is indicated when the objective specifies a required sequence of performances that is observed, and the performance is as important as the product. Process evaluation is also appropriate in cases where the product is not measurable for safety, or other constraining reasons.

Process evaluation is appropriate when one or more of the following is true:

- Detailed diagnostic information is desired
 - There is no product at the end of the process, or it is not measurable
 - The end product is not necessarily the result of the process alone (that is, perform life saving steps), or is dependent upon the actions of people or processes beyond the control of the learner
 - The product always follows from the process, but high costs, risks, or other practical constraints prevent evaluation of the product
-

Scoring Processes and Product Evaluations

Some situations may exist where both process and product evaluations are appropriate for a given objective. Some examples of conditions that may call for both product and process evaluation are:

- Although the product is more important than the process(es) which lead to its completion, there are critical steps, which, if not properly performed, may cause damage to equipment, or injury to personnel
 - The process and product are of similar importance, but do not assume that the product meets criterion levels
 - Diagnostic information is needed
 - Having process, as well as product measures, ensures information is obtained as to why the product does not meet the criterion
-

Product and Process Evaluation Roles

When obtaining both process and product measures for a specific objective, scoring must follow the criterion the objective specifies. That is, if the criterion specifies only a product, then do not use process scores to assess achievement of the criterion.

Designing Performance Test Items (continued)

Using Both Process and Product Evaluation

Three types of tasks illustrate the relative roles of product and process evaluation:

- Tasks where the product is the process. Relatively few tasks are this type. Singing a song or presenting a lesson are examples.
 - Tasks in which the product always follows from the process. These are tasks, such as fixed procedure tasks. If the process is correctly executed, the product follows. For example, following the correct process to repair a hydraulic system results in a repaired hydraulic system (the product).
 - Tasks in which the product may follow from the process. In a large number of tasks, the process appears correctly carried out, but the product not attained. Two reasons this happens is either the developer was unable to fully specify the necessary and sufficient steps in task performance, or did not accurately measure them. Weapon firing, for example, illustrates that there is no guarantee of acceptable marksmanship, even when following all procedures. In this case, process evaluation would not adequately substitute for product evaluation. Therefore, if any uncertainty exists, that using only a process measure may not adequately measure a student's ability to achieve the product or outcome of a task, add a product measure.
-

Preparing Checklist for Process and Product Evaluation

Performance Test Checklists (PCC)

A performance test checklist (PTC) is useful for rating ability to perform a specific set procedure. It is also a simple method of rating performance skills, when the purpose is to see if learners have reached a certain minimum level of performance.

For hard copy worksheets for performance test development, refer to Appendix N of SOP 5.

Unit 1: OS Fundamentals		Rev 01/14/2011					
PERFORMANCE TEST CHECKLIST (PTC)							
Name: _____				Class #: _____			
TPO 1.1							
Respond to an Electrical Shock emergency							
Student may use the following:		<input type="checkbox"/> Job Aid		<input type="checkbox"/> Reference Material		Other: _____	
EVALUATION CRITERIA:							
Accuracy:				Completeness:		Time:	
<ul style="list-style-type: none"> • Power correctly secured • Victim removed using nonconductive instrument • Victim placed in correct recovery position • No injury to self or others • Appropriate notification to command and/or supervisor 				Victim completely removed from source		Action is taken immediately	
STEP		Attempt					
		1 st		2 nd		3 rd	
		Y	N	Y	N	Y	N
Process Evaluation							
1. Secure Power		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Obtain help		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Remove victim from source		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"> ✓ Victim completely removed from source with nonconductive instrument ✓ Action taken immediately ✓ No injury to self or others 		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Assess condition of victim		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Brief responders		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Make notifications		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Product Evaluation							
1. Victim free of source		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"> ✓ Victim completely free, in correct recovery position 		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Response decisions:		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"> ✓ Accurate and immediate assessment of situation ✓ Accurate and immediate assessment of victim condition ✓ Correct decision – safe to assist without causing harm to self or others 		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Reports:		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"> ✓ Accurate and timely reports to command and/or supervisor 		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Additional notes for successful testing:		<input type="checkbox"/> Go <input type="checkbox"/> No Go		<input type="checkbox"/> Go <input type="checkbox"/> No Go		<input type="checkbox"/> Go <input type="checkbox"/> No Go	
Enter testing Scenario used							
Evaluator's Dated Initials:							
Course/School Chief dated Initials (required with 3rd attempt only)							

End of PT Checklist. To record Student progress, go to the Assessment Record on the following page.

Figure 7-1

Preparing Checklist for Process and Product Evaluation (continued)

Example Package of Performance Test

Refer to the *Performance Test Checklist Booklet* Appendix O of SOP Volume 5, Resident Instruction, for an example of a package developed for a performance test.

Developing a Process Checklist

When developing a checklist for rating, it is useful to distinguish process from product. The following is a list of steps to consider when developing a checklist for assessing the performance of a process:

Step	Action
1	Avoid lumping steps together. Example: If step #6 is not possibly accomplished without successful completion of the preceding steps, it may not require evaluation of those steps. Begin with the most logical step. Be careful not to mask a critical step when applying this technique.
2	Determine the time standard, or qualitative or quantitative standard, that applies. All or a portion of the elements may need evaluating in a process task.
3	Do NOT use a process test simply because the reference document shows a sequence. Ensure the sequence is critical, in terms of safety or security, to personnel or equipment.
4	Specify GO/NO GO criteria for each element measured.
5	Sequence steps in order of evaluation.
6	Keep performance steps statements short and precise.

Preparing Checklist for Process and Product Evaluation (continued)

Developing a Checklist for a Product

When preparing a checklist for scoring a product, it is useful to consider the following steps:

Step	Action
1	Specify the characteristics of the end product, prior to writing the test. Express standards in terms of (not all inclusive): <ul style="list-style-type: none"> • Shape • Strength • Time • Texture • Function • Tolerance • Tightness • Location • Consistency • Speed
2	Specify the product characteristics as performance elements.
3	Specify GO/NO GO criteria for each performance element and overall end product.
4	When preparing checklists for rating a combination process/product test, combine the appropriate construction rules from both test formats. Regardless of the type of test, ensure evaluation of all key elements (those critical to accomplishment, safety, or security).

Scoring Procedures for Performance Evaluations/Tests

Introduction

Criterion-reference testing lends itself to many forms of scoring. The purpose of a CRT is to differentiate masters from non-masters by comparing them to an absolute standard.

Types of PCC Scoring

A number of different types of test scoring procedures are available. Refer to a particular test, and consider the complexity of the tasks and/or products required, when choosing the proper scoring method. Some common types of Performance Criterion Checklist scoring include:

- Assist scoring
 - Pass/fail scoring
 - Fixed point systems
 - Rating scales
-

Assist vs. Non-interference Scoring

In testing, learners generally proceed from the beginning to end of a test, without comment or action on the part of the observer (noninterference). This type of scoring is often used in tests which call for the completion of a series of steps, or require production of a pre-specified product.

However, some tests may require scoring each step in a process. Thus, at each step, the learner's performance is approved (scored GO) or assistance is provided (and scored NO GO) before proceeding. Assist scoring is employed for diagnostic reasons. Remedial training then focuses on missed steps, saving retraining time and expenses. Assist scoring may also furnish valuable clues to areas where improvement in instruction is needed. For example, a large number of errors, in step number 3 of a 6-step procedure, may indicate a need for improved instruction in that area.

After preliminary training, a food service course objective might require testing a learner's ability to prepare a meal. Here, it is appropriate to observe each step in the planning, preparation, and serving of the meal, correcting and recording errors while observing. If the entire sequence is carried out properly, the product measure is scored GO. Errors observed may indicate the learner requires additional training on the deficient steps. Using an assist method of scoring not only obtains diagnostic information, but "saves" a meal—and the meal is served. Learners receiving assistance on the test receive a NO GO score; however, this minimizes the need for additional training before retesting. In this case, the test serves as another instructional event (or a practice with feedback) leading to mastery.

Scoring Procedures for Performance Evaluations/Tests (continued)

Pass-Fail Scoring

Generally, use noninterference scoring with performance tests. The simplest noninterference scoring is “pass-fail” scoring. It is generally used to score simple, objective “hard-skill” processes or products. Since the score is either “pass” or “fail,” the action is performed (or the product assembled or created) exactly as the objective specifies. The item is essentially an observable expression of the standard in the objective. Performance on the item either meets the standard, or it does not; there is no “gray” area. Examples of pass-fail scoring include:

- A trainee is given 10 minutes to detect and replace a defective transistor in a radio set. The trainee either does (pass) or does not (fail) get the unit operational within the allotted time.
- The Mount Captain must follow a specific procedure to activate the MK 75 Gun Mount System:

Step	Action
1	Positioned GCP switches and circuit breakers to the proper position.
2	Properly energized the hydraulic system.
3	Properly energized and tested the gun loading system.
4	Properly energized and tested the gun laying system.
5	Properly energized and tested barrel cooling (simulated).
6	Followed the procedure from memory without skipping steps
7	Followed the procedure from memory without skipping steps

Since this task is required in order for the MK 75 to be activated, a passing score is assigned only if no errors are observed on any of the items.

Fixed Point Scoring

Fixed point scoring is another type of CRT scoring.

This type of scoring is appropriate when the task or product scored is broken into several levels, which are quantitatively distinguished. For example, the item may call for adjusting values to specified tolerances. A trainee that adjusts them to the exact tolerance receives 4 points. Values adjusted to within $\pm .001$ inch = 3 points; $\pm .002$ inch = 2 points; $\pm .003$ = 1 point. No points are awarded if the trainee is off by $\pm .004$ of an inch or more.

Scoring Procedures for Performance Evaluations/Tests (continued)

Fixed Point Scoring (continued)

An alternate type of fixed point scoring uses “pass-fail” decisions on components of a task. For example, trainees are asked to overhaul a carburetor, and a point value assigned to different components of the task.

The following table provides an example of a pass-fail table, using a point value system.

Points	Task Description
1	Correct disassembly of carburetor.
1	Correct cleaning of carburetor.
1	Correct replacement of jets and parts of carburetor
1	Correct reinstallation of carburetor

A score of 4 indicates that all components of the task were correctly performed. If the trainee failed to replace the jets and float, but correctly performed components 1, 2, and 4, the score on the task, as a whole, is 3 points. A single test could test several tasks, each requiring performance on multiple components (subtasks).

Using a Checklist for Scoring

Scoring is generally performed using a checklist. All behaviors (or products) the objectives require are clearly defined. If the objective involves a product, scoring may compare the trainee’s product with a sample product. For example, if an objective requires filling, sanding, and painting a dented metal surface, to appropriate painting shop standards, each finished product (one painted surface) is compared to standard products. The top standard is a smooth, high-gloss metal surface. If the trainee’s product is similar to this, 4 points are awarded. The next standard is a smooth, high-gloss metal surface with slight ripples. If the trainee’s product resembles this, 3 points are given. This progresses down to the zero point standard—a metal surface finished so poorly that no points are assigned.

Scoring Procedures for Performance Evaluations/Tests (continued)

Cut-off Levels

Cut-off levels should reflect mastery of the objective, to the extent required. Since factors other than ability to perform a task (such as careless errors, evaluation errors, etc.) may affect an individual's score, cut-off levels are often set somewhat below 100 percent. If, for example, an objective calls for multiplication of two 4-digit numbers, the criterion might specify performing 10 such sets within 5 minutes, achieving the correct answer in at least 8 cases. Thus, the cut-off score of 8 (below 8 = fail) reflects an arbitrary definition of mastery. True mastery requires 10 out of 10.

Determining Mastery Cut-offs for Performance Tests

Although a properly constructed performance objective provides extensive data on what the performance standard is, it may not provide all that is needed. For example, it may not provide the number of successful iterations of performance necessary for mastery. (Is one enough?) Or, it may not indicate all the possible conditions under which the performance is tested. (Is one set of conditions enough?) Use SMEs to make judgments about what evidence is needed to assure mastery, such as:

- Number of iterations necessary as proof of mastery
 - Number of iterations necessary as proof of mastery
 - The scores to achieve (time, tolerances)
-

Consensus

The entire Training Development Team should participate in the mastery standard determination, but the opinions of the SMEs, with advice from the Instructional Systems Specialist, should predominate the decision-making process. The SMEs must agree on exactly what evidence to collect to determine mastery. If an agreement is not reached, include a third SME to arrive at a consensus.

Classifying Mastery

In making decisions about mastery standards, make the test design team acutely aware of the costs of:

- Wrongly classifying a master as a non-master (called a false negative classification), or;
 - Wrongly classifying a non-master as a master (called a false positive).
-

Scoring Procedures for Performance Evaluations/Tests (continued)

The Cost of Lowering Cut-off Scores to Meet Organizational Manpower Requirements

If the costs for false negatives are relatively high (e.g., manpower needs are critical) then lowering the cut-off score is justifiable. If the costs of false positives are high, then cut-off scores must remain high. In most cases, for training critical tasks, the costs to personnel and mission of a false positive (certifying non-master as master) are much more serious than certifying a master as a non-master. In the former case, certifying a non-master as a master may directly put personnel, equipment, or critical mission accomplishment at risk. In the latter case, the only result is likely just retraining to standard, or some delay in getting resources to the field.

Therefore, for most performance TPOs, the mastery standard is set very high to avoid certifying a non-master as a master. In any case, when performance on critical tasks is tested, cut-off points are kept high enough to reflect the standards specified in the objectives for those tasks.

Writing Performance Test Instructions

Instructions for PTC Participants

Once you are satisfied with procedures, directions, equipment, and scoring methods for a performance test, prepare detailed instructions to the test participants to formalize them. The primary requirement for test instructions is that they are complete, clear, and provided in writing. They should also include diagrams and pictures, as necessary, to fully describe the test environment to the instructor/administrator, the observer/evaluators, any actors, and the learner.

Refer to the *Performance Test Checklist Booklet* Appendix O of SOP Volume 5, Resident Instruction

Instructions to the Instructor and/or Test Administrator

Provide instructions to the test administrator in sufficient detail, so that an instructor who is competent in the area the test covers is able to set it up, run through the tasks, and then administer the test to learners, in a standardized way. These directions to the instructor and/or test administrator:

- Provide the precise procedures to follow
 - List the equipment needed
 - Point out especially hazardous aspects, or emphasize applicable safety precautions
 - Tell the instructor how to set up the equipment for the exercise
 - Define how the test is scored (in particular, gives details on whether assist or non-assist scoring is used)
 - State what questions from the learner, if any, the instructor/administrator can and cannot answer
-

Writing Performance Test Instructions (continued)

Directions to the Observers or Evaluators

In some instances, the administrator only administers the test; separate observers/evaluators (master task performers) may actually “grade” the learner. In this case, a set of instructions, different from the test administrator’s, is written for the observer/evaluator that precisely states their duties and responsibilities. These precise instructions indicate:

- What to observe (that is, performances expected to perform)
- How to record observations (that is, checklists, notes)
- How to compile and deliver observations
- Whether to discuss observation/ratings with other evaluators, to arrive at a consensus rating
- What guidance/instructions, if any, to give the learner
- Their role, versus the role of the administrator

Directions to the Learner

In very simple situations, you can give oral directions to the learners.

For example, to test an electrical technician’s ability, provide them with a wiring manual and avionics manual and ask them to troubleshoot a piece of equipment. However, such informality opens the door to the introduction of elements that could create non-standardized testing conditions. An instructor might give more detailed instructions to one individual than to another, or might inadvertently omit something important from the instructions.

To prevent such occurrences, read written instructions verbatim to the learner, or provide the learner the written instructions to study beforehand. The instructions should include (as necessary):

- Purpose of the test
- Time limits, if any
- Equipment provided
- Requirements the learner is expected to satisfy
- Special safety precautions
- Information on how the test is graded

Note: Certain situations may not require some of these items, but make a careful, informed judgment in each instance before omitting the information.

Writing Performance Test Instructions (continued)

Word of Caution on Instructions

Exercise care in developing the instructions, to avoid revealing unintended clues on proper procedure. Do not include any reference in the instructions that suggests a correct procedure on an earlier task, or provides correct results from previous procedures. An alert learner may take advantage of such unintended clues, resulting in an unfair advantage over other learners. This would attribute some of the differences in performance to “test-wiseness” or reading ability, rather than the ability to perform a given task.

For each performance TPO, develop a rating form (checklist), with directions for scoring. This form is highly individualized, specifying the checkpoints/steps on which the individual is evaluated. The determination of these checkpoints is vital. When determining the checkpoints, consider:

- Including as many as necessary, to ensure comprehensive coverage of the action indicated in the objective. Provide sufficient evidence that the learner can, or cannot, perform the action to the standard indicated in the objective.
 - Too few checkpoints probably indicate that some elements were overlooked.
 - On the other hand, too many checkpoints may suggest a failure to differentiate between critical and trivial elements.
 - The use of too many checkpoints may impose an impossible burden on the raters. It requires them to watch for too many things at one time and possibly miss the important factors while trying to grade performance on minor matters. For this reason, be selective and critical when developing the rating form.
 - Picking the items that are significant to successful performance. Choose items of a nature that are observed and judged with a high degree of objectivity.
 - Certain points may require the observer to check more than one item. For example, using a voltmeter in a physics project may require checking to ensure proper connection to the unit, and that the learner read the meter correctly. However, in many situations, you may desire the learner to record dial settings and meter readings on a separate form, specifically keyed to the instructions.
-

Writing Performance Test Instructions (continued)

Include All Safety Items as Checkpoints

In the checklist, include and evaluate all important steps, from a safety standpoint. For example, if there is a requirement to wear safety glasses while performing an operation, include a checkpoint such as: “Learner wearing safety glasses: GO _____ NO-GO _____. Do not allow learner to proceed with test until safety glasses are on.”

All or Nothing Scoring

The effectiveness of the evaluation process is reduced substantially, if the observer makes judgments about quality, along some sort of continuum. Experience shows that rating scales do not work very well in performance test situations. It is preferable to design the rating, at each checkpoint, on an “all or nothing” basis (that is, the learner did, or did not, do what was required; or, the evaluation was, or was not, correct within stated limits).

Chapter 8

ANALYTICS

Overview

Purpose of Test Analysis

Understanding how to interpret and use information based on candidates' or students' test scores is as important as knowing how to construct a well-designed test. Using feedback from the test is your guide to improving both the test and supporting instructional materials.

Performing analytics allows us to ensure the test items are both valid and reliable and are providing accurate measures of a learner's output. The analytics in this chapter focuses on multiple choice, single answer test items.

NOTE: The visual representations contained in this chapter demonstrate the concepts presented. Their presence is not meant to imply that the test developer must create these types of graphics when analyzing test results.

CRT vs. NRT

Interpretation of the scores and other statistics of norm-referenced tests (NRT) and criterion-referenced tests (CRT) are different.

- Criterion-referenced items are based on specific objectives that should be measured for mastery
 - Norm-referenced items will separate the scores of test takers from each other by rank ordering them.
-

Frequency Distribution

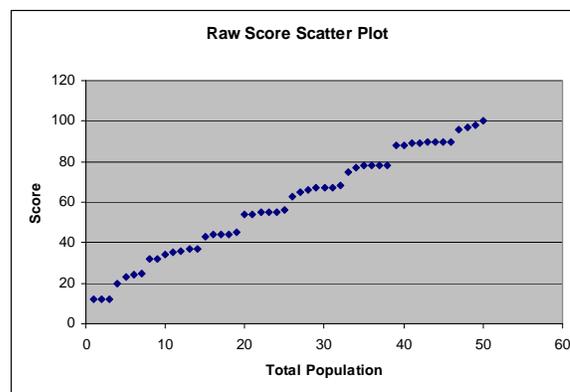
Raw Scores

A frequency distribution shows how many test takers achieved a specific score or fell within a defined range of scores. This can be numeric or it can be displayed graphically.

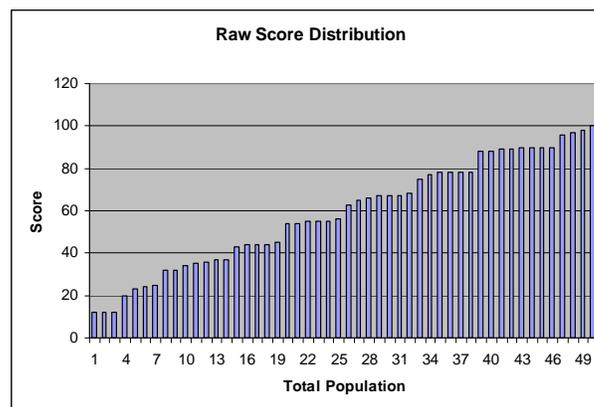
An example of a “perfect” raw score distribution is as follows:

- If 100 candidates take a 100 question norm-referenced test, the scores would range from 0 to 100 with each test taker achieving each one of the scores; 0, 1, 2, 3, 4, 5, etc.

In the frequency distribution of individual raw scores below, the scores versus the number of test takers is displayed in a “scatter plot.” In the chart below, three test takers scored between 0 and 20. Twelve test takers scored between 80 and 100, etc.



The same data can also be displayed in the more familiar bar chart.

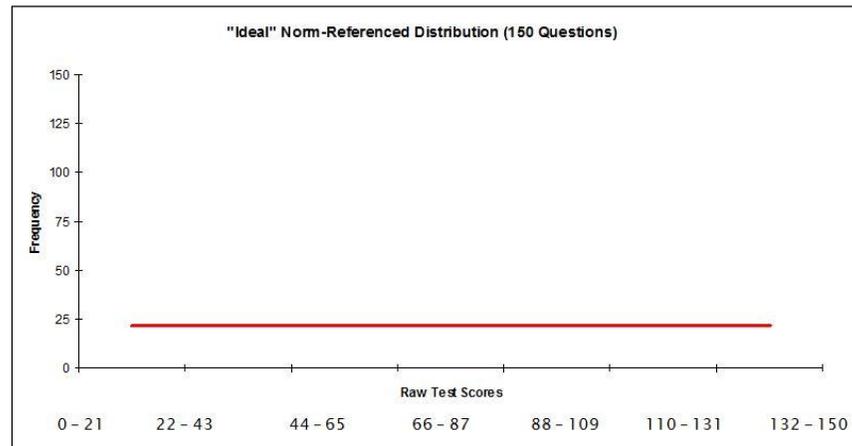


Frequency Distribution (continued)

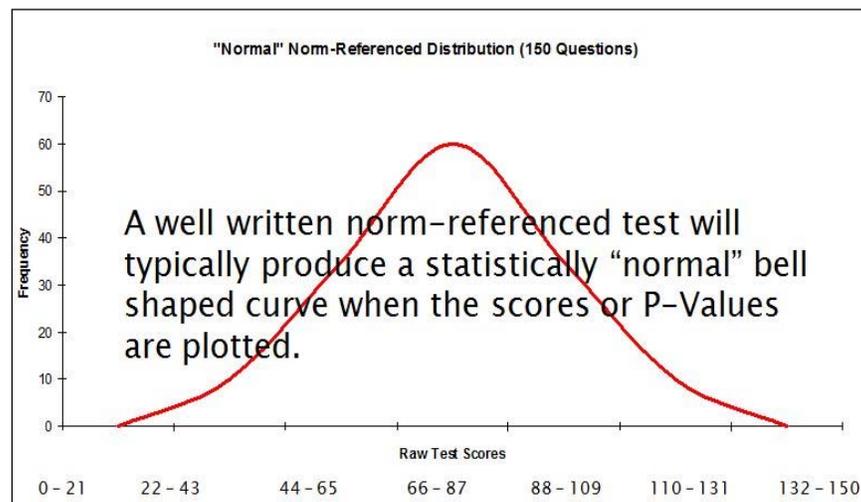
NRT Normal Curve

Plotting the results of a norm-referenced test by frequency and raw test scores usually results in a “normal” bell-shaped curve.

Ideally, the plot for a norm-referenced test should be a flat line and not a “curve”.



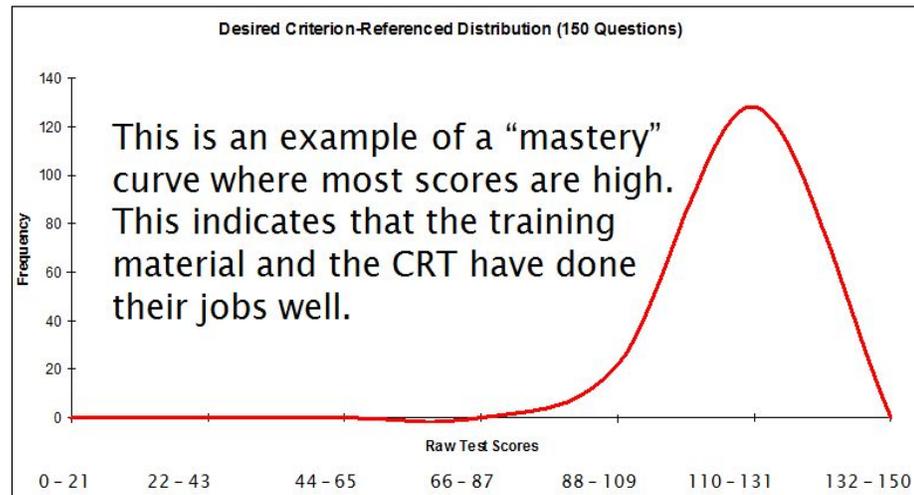
The purpose of a norm-referenced test is to rank-order personnel. The more spread out across the range of scores, the better job the norm-referenced test has done. If scores are clustered towards either end (skewed), the test is not discriminating.



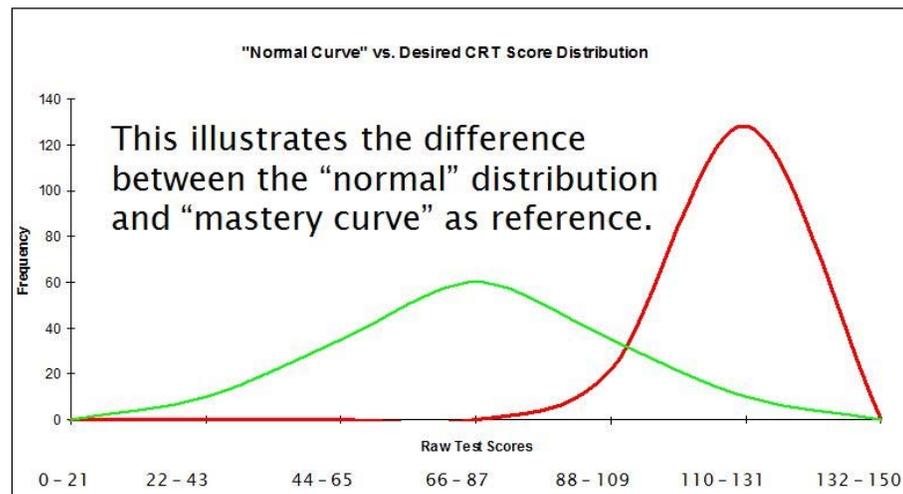
Frequency Distribution (continued)

CRT Distribution Curve

Unlike the NRT which compares one candidate to another, the purpose of a criterion-referenced test is to assess whether a learner has mastered the content. When test takers raw scores are plotted, it is desired that the curve is skewed left (left tail longer) as shown below.



There is no limit as to how many candidates can succeed on a CRT. The test is considered a test of mastery of a concept or set of concepts. Technically, there is no “normal” distribution but it may be used as a reference point for visualization of the results as shown below.



Commonly Reported Statistics

Introduction

Aside from frequency distributions, there are a number of useful testing statistics that are commonly examined.

Mean

The mean of a set of test results is the average score. The mean can be expressed as a raw number or by a percentage. It is calculated by adding all scores and dividing by number of test takers. For example, for the scores 30, 45, 67, 90, 77, 36, 97, 65, 84, 79, the total is 670 and the mean is 670 divided by the total number of scores, 10. The mean of this set of scores is 67. ($670 \div 10 = 67$)

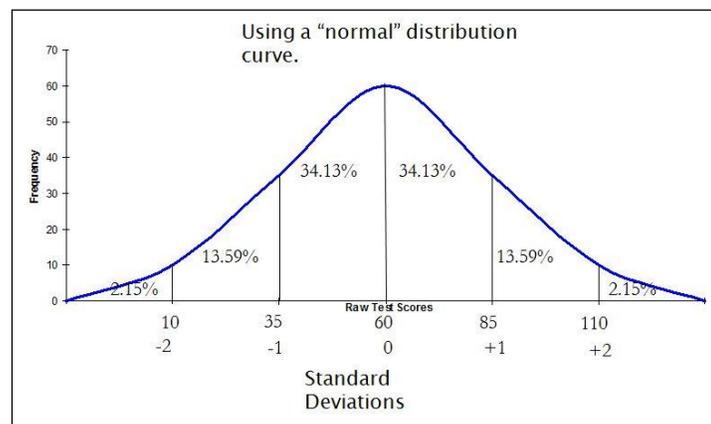
Standard Deviation

The standard deviation is a function of the bell curve that defines the average deviation or degree of distribution of scores from the mean score. It tells you how far from the mean score a percentage of test takers deviate. For example, for the set of scores mentioned above in the description of the term mean (30, 45, 67, 90, 77, 36, 97, 65, 84, 79), the standard deviation is approximately 23.

This means that a certain percentage (34.13%) of the population scored within 23 points, above or below the mean (1 SD), 13.59 % were ± 46 points of the mean (2 SD), and 2.15% were ± 69 points of the mean (3 SD) and 0.13% were ± 92 points (more than 3 SD) or more from the mean.

A large standard deviation indicates that the data points are far from the mean and a small standard deviation indicates that they are clustered closely around the mean.

The graphic below depicts the standard deviation for a 150 item test with a SD of 25:



Commonly Reported Statistics (continued)

Acceptable Standard Deviation

For a NRT, the acceptable range for a standard deviation is 12.00 to 18.00. If the standard deviation is low for a norm referenced test, too many candidates are getting scores close together, which means the test is not serving the purpose of discriminating test takers from each other by test score.

For a CRT there is no specific “acceptable” range for the SD. Typically the SD should be small when the candidates are grouped together which shows that their scores do not deviate greatly and the majority of them achieved much the same score. They mastered the concept(s).

Reliability

Reliability is the extent to which the test is effective at measuring anything at all. A reliable test is one which has the capacity to repeat the same statistical results repeatedly. A test with excellent reliability is one with 0.90, or 90%, or higher, probability of repeating the same statistical results. The goal of the USCG is to maintain test reliability at or above .80. Table 8-1, which was adapted from SCOREPAC[®] Item Analysis, is an interpretation of reliability values.

Reliability	Interpretation
.90 and above	Excellent reliability
.80 - .89	Very good (USCG recognizes this as acceptable)
.70 - .79	Good – range of most. Some items could be improved
.60 - .69	Somewhat low. Test should be supplemented by other measures to determine performance. Some items need improvement.
.50 - .59	Suggests need for revision of test. Test definitely needs to be supplemented by other methods.
Below .50	Questionable reliability. The test should not contribute to the grade and needs revision.

Table 8-1

Commonly Reported Statistics (continued)

P-Value

The P-value is also called the difficulty index. There are different types of P-values. The correct response P-value is the measure of the number of people who answered a specific item correctly. A mastery criterion-referenced test should have a high difficulty index of .80 or higher. The P-value, or difficulty index, of a norm-referenced test should range from .28 to .80.

Another P-value is derived from the test takers who selected each of the incorrect options. This is called the incorrect response P-value. The correct response P-value should be higher than each of the incorrect P-values. In other words, the correct response should be chosen more than any of the distractors, even in difficult questions. The P-value can be used to evaluate the validity of a test item and to evaluate the validity of each distractor. The distribution of incorrect response P-values should be relatively equal if all of the distractors (alternatives) are well written.

The P-value of all test items should be reviewed periodically to ensure the test items are performing as expected. The P-value alone can tell you quite a bit about the test item but must be paired with the *d*-value to present a full picture.

d-Value

The *d*-value represents the statistical function, “Point-Biserial Correlation,” commonly called the discrimination index. This may sometimes be identified as the “alt *r*-value.” It is the degree to which the test item differentiates between those who know the material well and those who do not. Like the P-value, the discrimination index evaluates the choices that the test takers make.

The correct response *d*-value should always be a positive number and the incorrect responses *d*-value should always be negative. The range of the *d*-value is negative 1.000 to positive 1.000.

- Good correct item *d*-value range is +0.250 to +0.750
- Good incorrect item *d*-value range is -0.250 to -0.750

A large positive *d*-value such as .40 for the correct answer means that test takers with high scores on the test are also getting the item correct and those with lower test scores are getting the item wrong. A low *d*-value for the correct response implies that test takers who get the item correct tend to do poorly on the exam overall and those who do well on the exam tend to get the item wrong. A negative *d*-value for a correct response indicates there is some deficiency in the item which may include: item keyed incorrectly, item poorly constructed, misleading distractors, content inadequately taught, etc.

Commonly Reported Statistics (continued)

General Discrimination Range

Table 8-2 provides an interpretation of item quality based upon the d -value. Note the values in the table are absolute values, correct choices should be positive and incorrect choices should be negative.

Absolute Value Range	Item Quality
0.50 or higher	Very high discrimination
0.30 to 0.49	High discrimination
0.16 to 0.29	Moderate discrimination
0.15 or less	Low discrimination Review item to determine reason
Note: Negative values are expected for incorrect responses. If a correct response has a negative value a problem is indicated.	
Adapted from: Pope, G. 2009, <i>Item analysis analytics</i> . Questionmark Corporation. Retrieved January 17, 2013, from http://www.questionmark.com/us/whitepapers/index.aspx	

Table 8-2

P & d -Value Generalizations

Generalizations can be made by observing the relationship between the P and d values.

- Very easy or very difficult test items have very little discrimination value – they do not tend to separate test takers who fully understand the material and those who don't.
 - Items of moderate difficulty (60% to 70% answering correctly) generally are more discriminating.
 - If all test takers respond correctly or incorrectly the item does not discriminate at all on a NRT and probably should be removed.
 - If all test takers respond incorrectly on a CRT the item may be keyed wrong or being taught inadequately.
-

Commonly Reported Statistics (continued)

P & *d*-value Interpretation

Table 8-3 presents sample P and *d*-values for five test items with commentary describing the interpretation. Correct responses are indicated by **bold underline**.

Item Number	Response "A"		Response "B"		Response "C"		Response "D"	
	P	<i>d</i>	P	<i>d</i>	P	<i>d</i>	P	<i>d</i>
1	<u>.0396</u>	<u>0.261</u>	0.271	-0.207	0.208	-0.112	0.125	0.031
2	<u>0.396</u>	<u>0.370</u>	0.000	0.000	0.083	-0.186	0.521	0.629
3	0.208	0.260	<u>0.208</u>	<u>0.030</u>	0.063	-0.207	0.521	-0.135
4	0.000	0.000	0.000	0.000	0.000	0.000	<u>1.000</u>	<u>0.000</u>
5	0.417	0.061	0.125	-0.214	0.271	0.231	<u>0.188</u>	<u>-0.159</u>

Table 8-3

- Item #1** – The P-value indicates 39.6% of test takers selected the correct response "A." Alternatives "B" & "C" were selected pretty much equally while "D" was only selected by 12.5% of test takers – "D" may be a weak distractor. The correct response *d*-value is positive and shows moderate items discrimination. "B" and "C" *d*-values are negative as they should be. "B" has moderate discrimination but "C" is low and should be evaluated. Response "D" has a weak positive discrimination which correlates with the low percentage of test takers selecting the response. Response "D" may be misleading.
- Item #2** – The correct response is keyed as "A" but the P-value indicates 52.1% of test takers chose "D". Very few chose "C" and no one chose "B". This item may be keyed wrong and appears to be a poorly written or confusing item. The *d*-value of response "D" confirms that the item may be keyed incorrectly and "D" is the actual correct response. That should be the first consideration. If the item was keyed incorrectly the *d*-values will change based on overall test performance. Overall – the item is a poor item and should be reviewed.

Commonly Reported Statistics (continued)

P & *d*-value Interpretation (continued)

- **Item #3** – The correct response is keyed as “B” but 52.1% of test takers chose “D”. Responses “A” & “B” were chosen equally while “C” was chosen by only 6.3%. This item may be keyed wrong or distractor “D” may be confusing or misleading. The *d*-value for response “A” is a moderate positive and it should be negative which means the test takers who know the material are being led to respond more frequently to “A” rather than the correct answer. The item should be reviewed.
- **Item #4** – The “P” value indicates all test takers responded correctly. If this item appears in a NRT it is not discriminating or sorting out the most from the least knowledgeable and should be considered for removal. If it is on a CRT it confirms that all of the test takers have mastered the concept but should still be reviewed to ensure the distractors are not spurious. If subsequent results continue to be 1.00 for every administration of the test the item should be considered for revision or removal because it is not providing any discrimination.
- **Item #5** – This item is either keyed incorrectly or very misleading because most test takers (41.7%) are selecting “A” while only 18.8% are selecting “D”. The moderate negative *d*-value for “D” is another indicator that something is amiss as are the positive *d*-values for “A” and “C” which should be negative.

Item Total Outcome Correlation

Item-total outcome correlation is a point-biserial calculation that compares a test item’s score with the test taker’s total exam score. Higher item scores should mean higher exam scores overall. A high item-total correlation represents a higher internal test consistency and reliability. It means that test takers that score high on the test also scored higher on the test item than test takers that scored low on the test. A low item-total correlation means that the test takers who scored low on the test are getting the answer correct more often than the test takers who scored high on the test. You would want to take a look at a test item that has a low item-total correlation because it seems to be confusing test takers who are demonstrating high competence. The total-item discrimination is calculated using statistics for the test takers who score in the upper 27% minus statistics for those who score in the lower 27%. There should be a large positive difference between low and high.

Commonly Reported Statistics (continued)

Interpretation of Total Outcome Correlation

Table 8-4 provides interpretation of the total outcome correlation range.

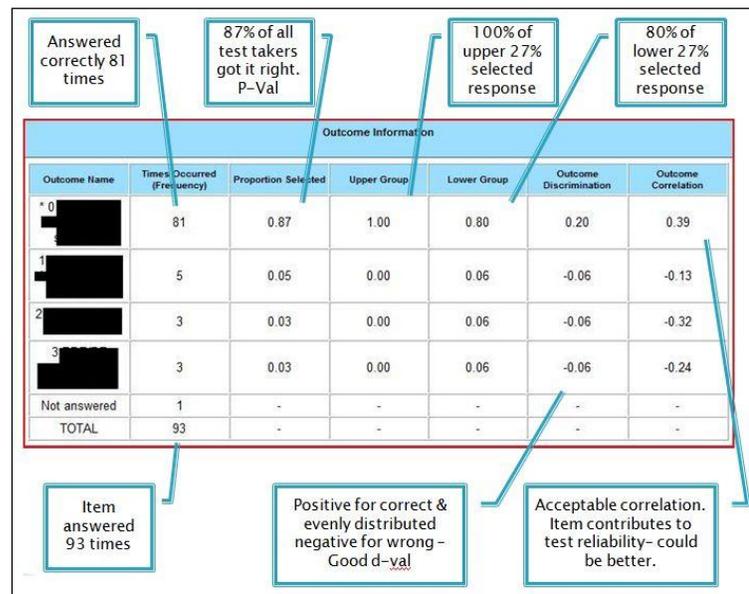
Total Outcome Correlation	Interpretation
Negative	Major problem indicated if this is occurring for a correct distractor – find out why.
Around zero	No relationship between the test item score and the total assessment score. Review the items to determine why.
0 – 0.19	Low correlation between outcome scores and assessment scores.
0.20 to 0.29	Moderate correlation between outcome scores and assessment scores.
0.30 to 0.44	Strong correlation between outcome scores and assessment scores.
0.45 or greater	Very strong correlation between outcome scores and assessment scores.

Pope, G. 2009, *Item analysis analytics*. Questionmark Corporation. Retrieved January 17, 2013, from <http://www.questionmark.com/us/whitepapers/index.aspx>

Table 8-4

Item Outcome Report

The illustration below represents an item outcome from the current USCG testing system, Questionmark Perception.

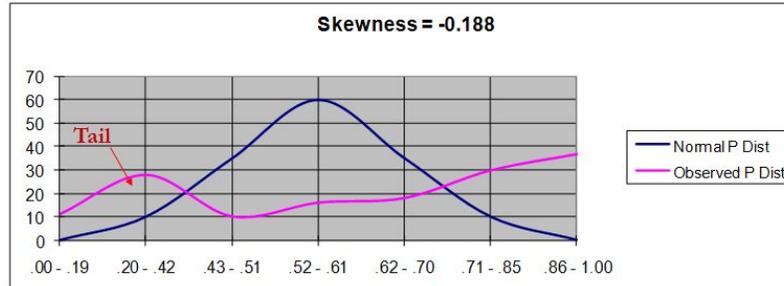


Commonly Reported Statistics (continued)

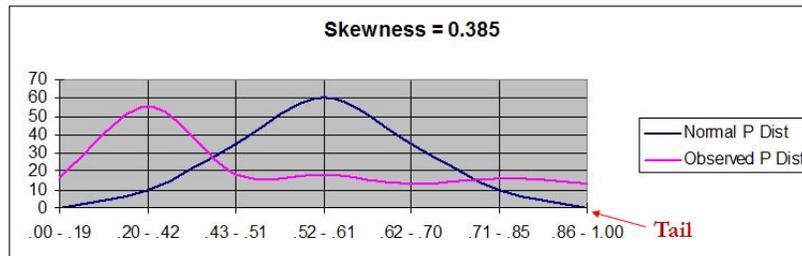
Skewness

Skewness refers to how the data looks when plotted on a graph.

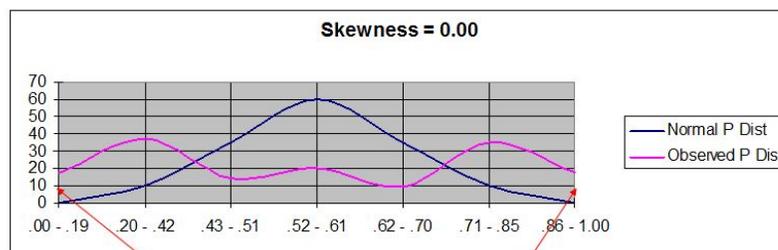
Negative values usually indicate a relatively easy test. A negative skewness is said to be “skewed left” which means the left “tail” is longer relative to the right “tail” as illustrated below.



Positive values usually indicate a difficult test. A positive value is “skewed right” in that the right tail is longer relative to the left tail as shown below.



A zero value may indicate a “normal” distribution but must be plotted to verify this. In the illustration below the skewness is zero but it is actually a two tailed distribution. The results shown are from a NRT so overall the distribution is “normal” in that the observed results are almost “flat.”



Two-Tailed Distribution

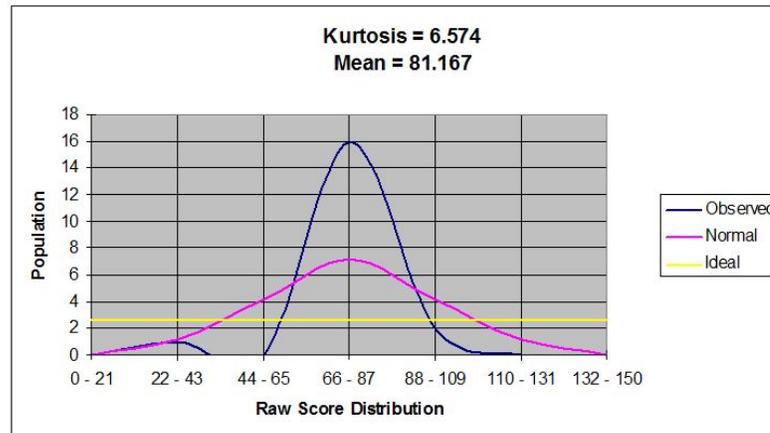
Commonly Reported Statistics (continued)

Kurtosis

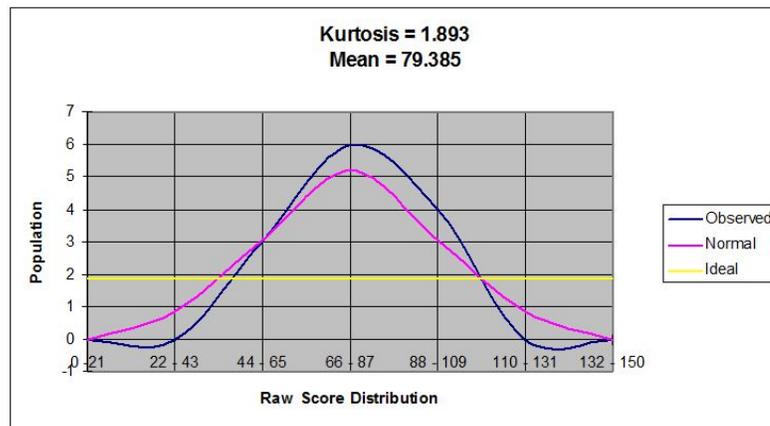
Kurtosis is a measure of “peakness” of a distribution. Another way to view this is flatness opposed to pointed when compared to a “normal” distribution curve.

A “normal” kurtosis, which is very rare, will have a value of 0.00.

A high kurtosis value indicates a distinct peak near the mean that declines rapidly and has a heavy tail. This is common in a CRT but not desired in an NRT.



A low kurtosis value indicates a relatively flat top near the mean as shown below. This is desired in an NRT.



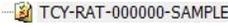
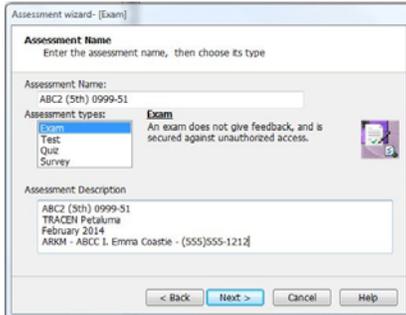
If the Kurtosis value is not as expected, depending upon the goal and type of test, the test items should be adjusted in attempt to correct the outcome. For an NRT the item difficulty should be adjusted upward to force the dataset to spread out. The items in a CRT should be reviewed to ensure they are valid and subject matter is being presented correctly.

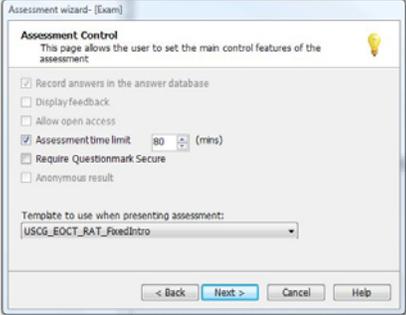
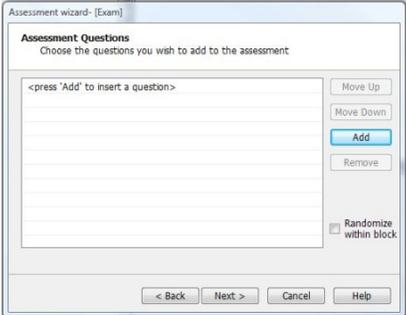
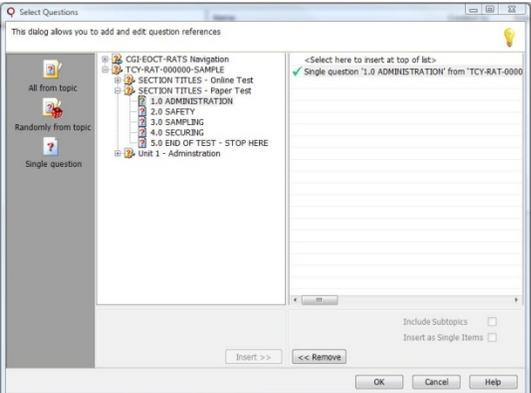
THIS PAGE LEFT INTENTIONALLY BLANK

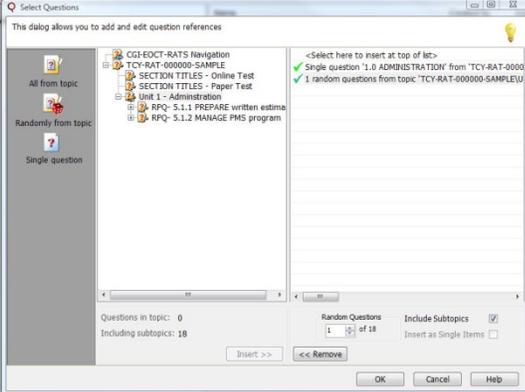
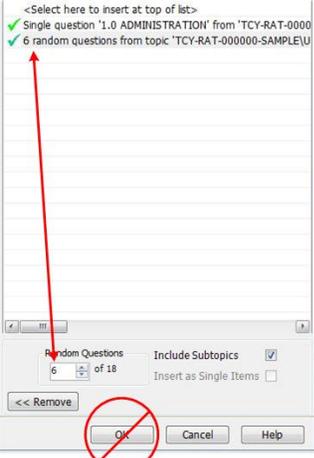
Appendix A

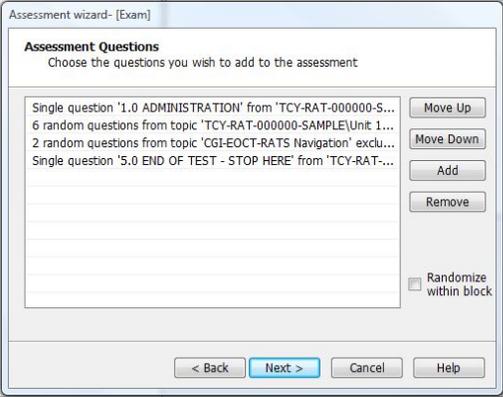
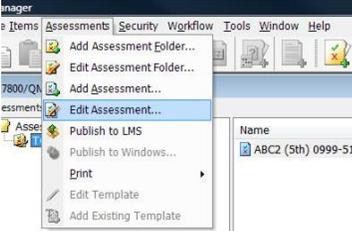
JOB AID FOR TEST DESIGN OF THE PAPER VERSIONS OF THE RAT USING QUESTIONMARK PERCEPTION

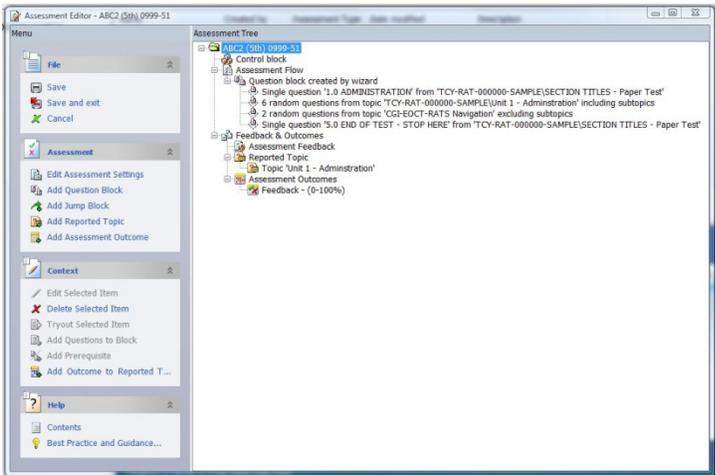
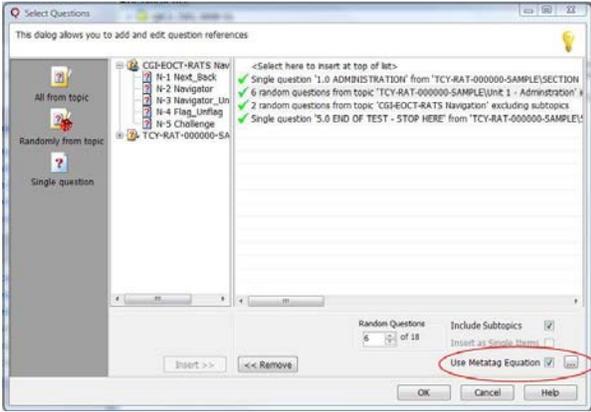
Job Aid

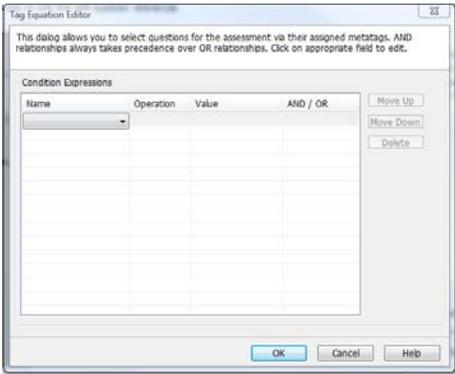
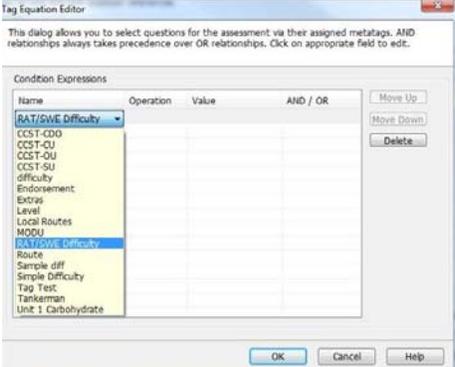
Step	Action
1	From within Authoring Manager, select <i>Assessments view</i> icon.  the
2	Left click on appropriate RAT folder.  assessment
3	Left click on the <i>Add Assessment</i> icon in toolbar. 
4	The <i>Assessment Wizard</i> will open. Left click on <i>Next</i> to continue. The <i>Assessment Name</i> screen will open.
5	<p>Enter a name for the RAT. The name that is entered here is what will appear on the RAT, either the paper or the electronic version. Use the naming conventions illustrated here and described in the RAT Naming Convention section in this chapter.</p> <p>Note: It is recommended that the paper version be designed first, then copied and modified for the online version.</p> <div style="display: flex; justify-content: space-between;"> <div style="width: 40%;"> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> MST2 (6th) 0234-51 <input checked="" type="checkbox"/> MST2 (6th) 0234-52 <input checked="" type="checkbox"/> MST2 (6th) 0234-53 <input checked="" type="checkbox"/> MST2 Rating Advancement Test 0234 (6th) </div> <div style="width: 55%; text-align: right;"> <p>recommended designed first, modified for</p> </div> </div>
6	<p>Ensure “<i>Exam</i>” is highlighted.</p> <p>Enter the following “Assessment block.”</p> <ul style="list-style-type: none"> - Assessment Name, - TRACEN responsible - Effective Date - RKM/ARKM contact <div style="display: flex; justify-content: space-between; align-items: center;"> <div style="width: 40%;"> <p>in the Description”</p> <p>Edition, Rate</p> <p>information</p> </div> <div style="width: 55%; text-align: center;">  </div> </div>
7	Left click <i>Next</i> .

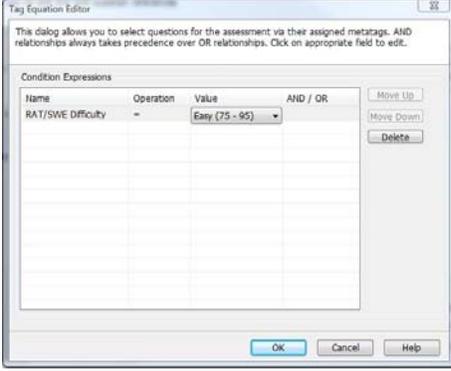
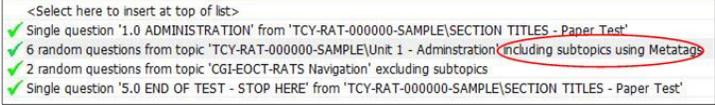
Step	Action
8	<p>The Assessment Control window will open. This is where time limits are set, whether Questionmark Secure will be required, and what template to use are selected. These settings do not affect the paper test but may be assigned to save time when copying the test for electronic display.</p> <p>Check Assessment time limit and fill in the block for the time limit of the assessment – unless otherwise directed by the RTAC this will be two minutes for each test item. (Example: 40 item test = 80 mins.)</p> <p>Require Questionmark Secure will only be checked for the electronic version after consultation with the Non-resident Training Branch at Coast Guard Institute.</p> <p>Set the Template to USCG EOCT RAT FixedIntro by selecting from the drop down menu.</p> 
9	Left click Next .
10	Assessment Introduction will remain blank.
11	Left click Next .
12	<p>The Assessment Questions screen will appear. This is where you begin the process of test item (question) selection for the RAT.</p> 
13	Left click Add .
14	<p>The Select Questions wizard screen will open. Left click the “+” sign next to the SECTION TITLES to expand the selections.</p> <p>Left click Single Question.</p> <p>Left click the section title you will be adding.</p> <p>Click Insert.</p> 
15	The selected Single Question which was previously created in the SECTION TITLES subtopic will appear at the top of the list as shown in step 14. Left click Randomly from Topic .

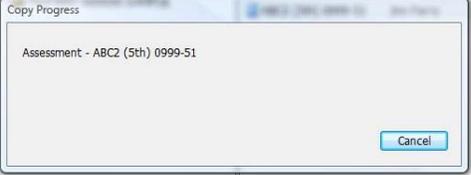
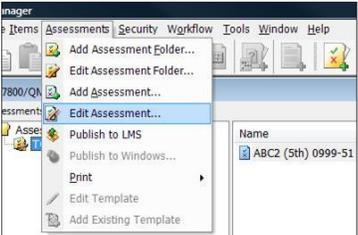
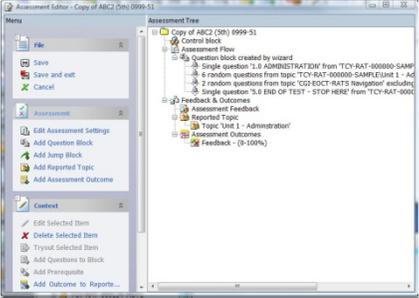
Step	Action
16	<p>Left click the unit you wish to add items from. Left click Insert.</p> <p>You will notice the Wizard screen automatically adds “1 random question[s] from topic ‘...’ and opens a box at the bottom titled Random Questions with a number selection box. The box shows the number of questions available in the selected topic.</p> 
17	<p>Enter the desired number here as defined in the test design spreadsheet. Ensure the Include Subtopics box is checked. Left click on the question list and it will change to the number you have selected.</p> <p>In this example we have selected six (6) random questions out of 18 available.</p> <p>Note: Do NOT click OK or you will have to reopen the dialog box and edit the assessment to continue.</p> 
18	<p>Continue adding test items in this manner until the test is populated with the correct number of items from each unit/topic to match the requirements of the test design spreadsheet.</p> <p>Note: To ensure RATs of equal difficulty, you must enter the random number of items desired from each unit to match the number of Easy, Moderate and Hard items calculated.</p> <p>Example: If 6 Easy and 2 Moderate items are desired from Unit 1 the selection must be made twice from unit 1. The metatags defining difficulty will be added later.</p> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p><Select here to insert at top of list></p> <ul style="list-style-type: none"> ✓ Single question '1.0 ADMINISTRATION' from 'TCY-RAT-000000-SAMPLE\SECTION ✓ 6 random questions from topic 'TCY-RAT-000000-SAMPLE\Unit 1 - Administration' i ✓ 2 random questions from topic 'CGI-EOCT-RATS Navigation' excluding subtopics ✓ Single question '5.0 END OF TEST - STOP HERE' from 'TCY-RAT-000000-SAMPLE\ </div>

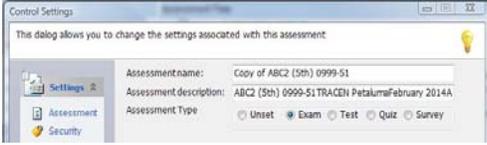
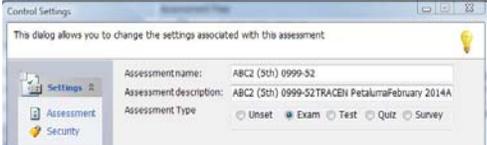
Step	Action
19	<p>When all of the items are selected, left click OK. The Assessment Questions summary screen will appear. Review this screen to ensure your test is designed as desired. If edits are necessary, select the appropriate buttons.</p> <p>Ensure Randomize within block is NOT checked. If the items are randomized within the block, the section titles will also be randomized. When the RAT is printed, the items will be randomized automatically.</p> 
20	Left click Next .
21	The Assessment Feedback screen will appear. Left click Next – you will not provide feedback for this assessment.
22	Left click Finish .
23	<p>The assessment will appear in the list box for the assessment folder.</p> 
24	<p>You will now add the metatag routine to select the correct number of items from each difficulty level.</p> <p>Left click the RAT you wish to edit to highlight it.</p>
25	<p>Left click Assessments on the toolbar to open the Assessments drop box.</p> <p>Left click Edit Assessment.</p> 

Step	Action
26	<p>The assessment will open in editor view.</p> 
27	<p>Left click on the first set of random questions chosen from Unit 1 to highlight it. Left click on <i>Edit Selected Item</i> on the left to open the editor. When the box opens, highlight the set of items again. The <i>Use Metatag Equation</i> checkbox will become active.</p> <p>Place a check in this box by left clicking.</p>  <p>Left click the box with three dots “...” to the right of <i>Use Metatag Equation</i>.</p>

Step	Action
28	<p>The Tag Equation Editor box will now open.</p>  <p>This is where you will assign the difficulty level to be set for each set of test items.</p>
29	<p>In this example, we are assigning a difficulty level of Easy to the first six random test items.</p> <p>Left click on the dropdown beneath Name in the Condition Expressions box.</p> <p>Select RAT/SWE Difficulty from the choices available.</p> <p>Note: Choices available may vary depending upon your level of access and system administrative changes.</p> 
30	<p>Left click on Operation and select “=”.</p>

Step	Action
31	<p>Left click on <i>Value</i> and select “Easy (75 – 95)”.</p>  <p>Note: The other choices are “Hard (25 – 49)” and “Moderate (50 – 74)”. Depending upon the difficulty desired for the set of test items your selections will vary.</p>
32	Left click OK .
33	<p>If you expand the question list, you will notice the phrase ‘...including subtopics using Metatags’ has been added to the selected set of test items.</p> 
34	Continue adding metatags to correspond to the desired RAT design until completed. When you are satisfied with the design, left click OK .
35	<p>Review the Reported Topic list and delete those that will not be required to be reported on for analytics. These include SECTION TITLES, CGI-EOCT-RATS Navigation and usually all sub-topics within each unit.</p>  <p>Note: You may want to keep sub-topics within each unit should you foresee a reason to perform analytics at a more detailed level.</p> <p>When you are satisfied with your RAT design, be sure to click Save and exit!</p>

Step	Action
36	<p>To save time and ensure the remaining two parallel versions of the RAT are of the same design and difficulty; highlight the RAT that you just built by left clicking on it.</p> <p>Select Edit from the top menu bar and left click <i>Copy</i>. A <i>Copy Progress</i> box will appear.</p> 
37	<p>When the <i>Copy Progress</i> box disappears, move your cursor to the list of tests and click in the open area.</p> <p>Move your cursor <i>Edit</i>, left click and select <i>Paste</i>.</p> <p>Note: Depending upon length and complexity of the RAT as well as the number of items available in the database this step may take some time.</p> <p>The <i>Paste Progress</i> box will appear briefly. A copy of the original RAT will be added to the list of assessments.</p>
38	<p>Repeat steps 36 and 37 to make the third copy.</p> 
39	<p>You must now edit the copies to assign a new name that corresponds to test 52 and 53 (or other versions). Start by selecting the first copy of the original assessment in the list.</p>
40	<p>Left click <i>Assessments</i> on the toolbar to open the Assessments dropdown box.</p> <p>Left click <i>Edit Assessment</i>.</p> 
41	<p>The copy of the original assessment will open in the <i>Assessment Editor</i> and show the <i>Assessment Tree</i>.</p> 

Step	Action
42	<p>Left click on <i>Edit Assessment Settings</i>.</p> <p>The <i>Control Settings</i> screen will open.</p> <p>Edit the Assessment name and Assessment description to the desired name.</p> <p>Example: Change “Copy of ABC2 (5th) 0999-51 to ABC2 (5th) 0999-52.</p>  <p>Is changed to:</p> 

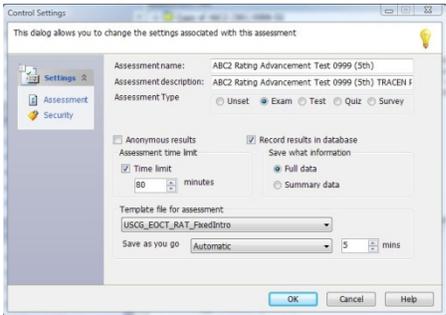
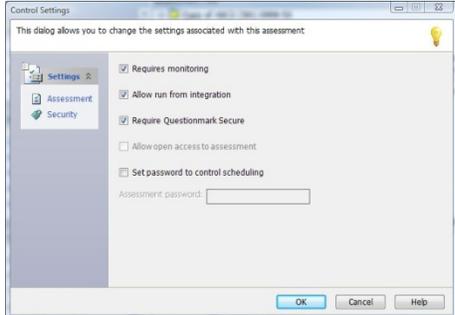
THIS PAGE LEFT INTENTIONALLY BLANK

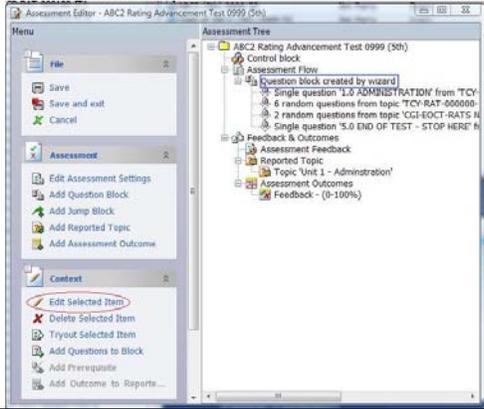
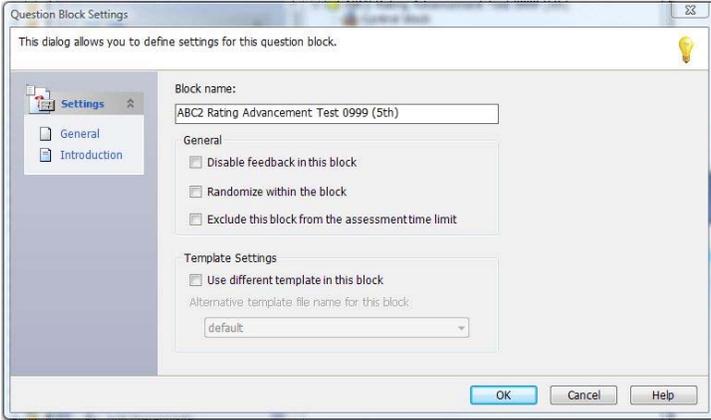
Appendix B

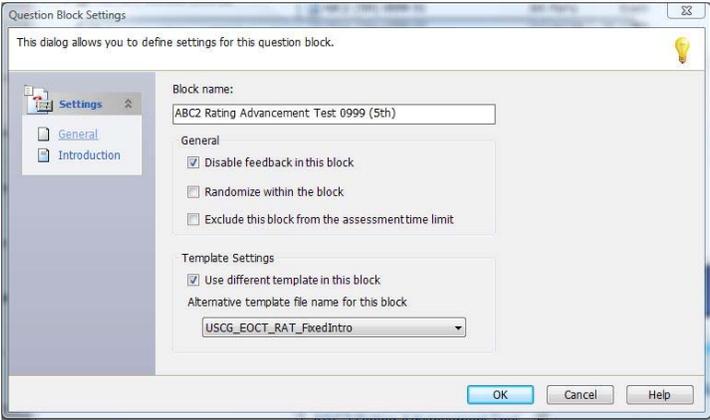
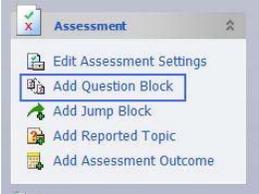
JOB AID FOR TEST DESIGN OF THE ONLINE VERSION OF THE RAT USING QUESTIONMARK PERCEPTION

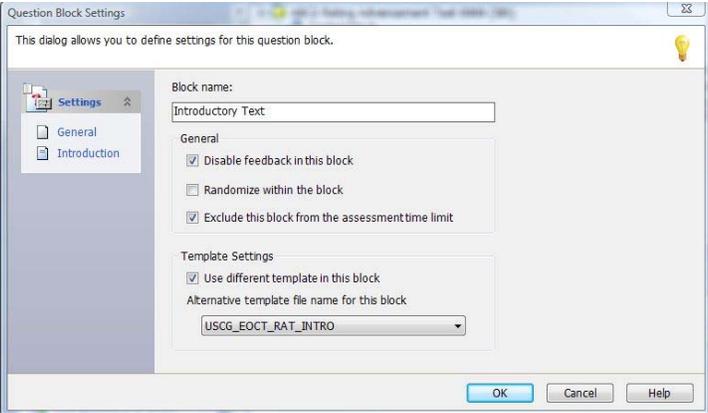
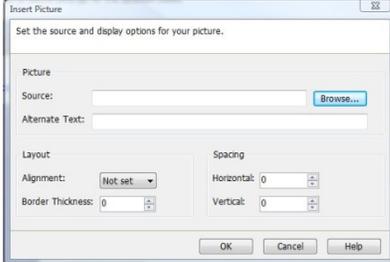
Job Aid

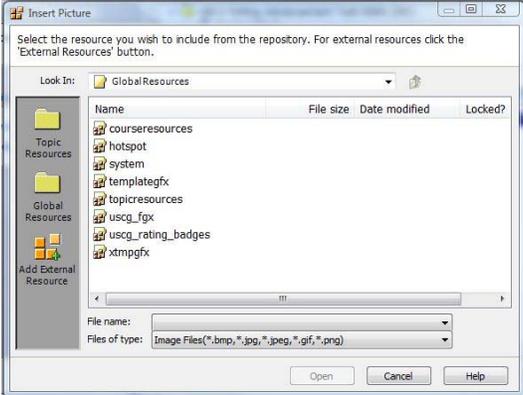
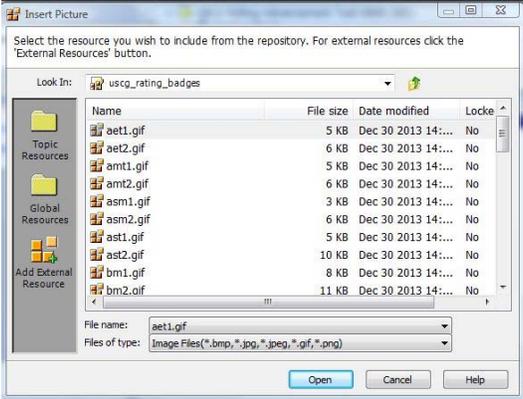
The following Step Action table describes the process of constructing the online version of the RAT.

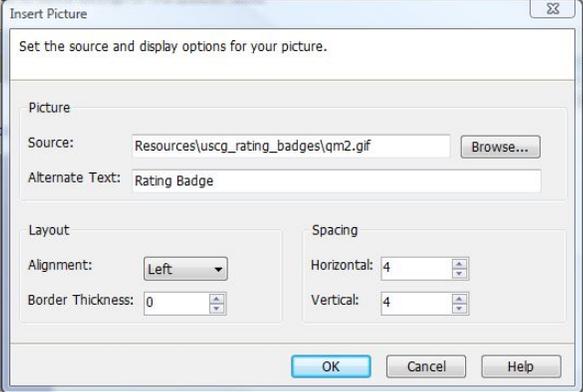
Step	Action
1	Make a copy of an existing paper version of the RAT following the procedures in the step action table for RAT “Construction From Within Questionmark Perception.”
2	Open the Control Settings dialog box by selecting Edit assessment . Edit the copy to reflect the online RAT name. “ABC2 Rating Advancement Test 0999 (5 th)”
3	<p>While the Control Settings dialog box is open ensure the following additional settings are as desired:</p> <p>Assessment Type • Exam</p> <p>√ Record results in Save what information</p> <p>• Full data</p> <p>√ Time Limit [# of test items times 2]</p> <p>Template file for assessment</p> <p>USCG_EOCT_RAT_FixedIntro</p> <p>Save as you go Automatic [5] mins</p>
	 <p>database</p>
4	Left click Security.
5	<p>Ensure the following are checked:</p> <p>√ Requires monitoring</p> <p>√ Allow run from</p> <p>√ Require Questionmark otherwise directed by CGI)</p>
	 <p>following integration Secure (unless</p>
6	Click OK .

Step	Action
7	<p>From the Assessment the Edit highlight “Question block by wizard” and “Edit Selected”</p>  <p>Tree in window, created left click Item”.</p>
8	<p>Change the Block name from Question block created by wizard to match the name of the RAT.</p> 
9	<p>Ensure Disable feedback in this block is checked and the other choices in General remain unchecked!</p>

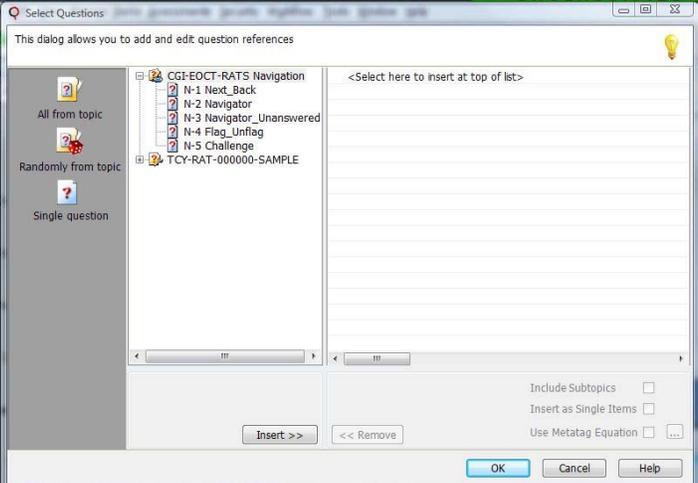
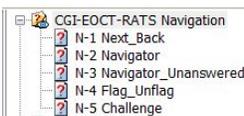
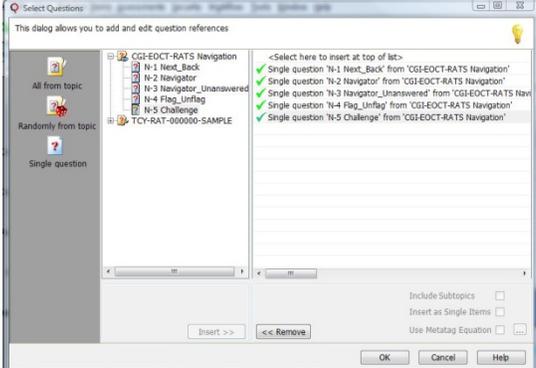
Step	Action
10	<p>Under Template Settings, check the box: \checkmark <i>Use different template in this block</i></p> <p>The <i>Alternative template file name for this block</i> will darken and become active.</p> <p>Select <i>USCG_EOCT_RAT_FixedIntro</i> from the dropdown box.</p> 
11	Click OK
12	<p>You will now create the Introductory Question Block.</p> <p>Left click Add Question Block Menu.</p> <p>The Question Block Settings will open.</p>  <p>from the dialog box</p>

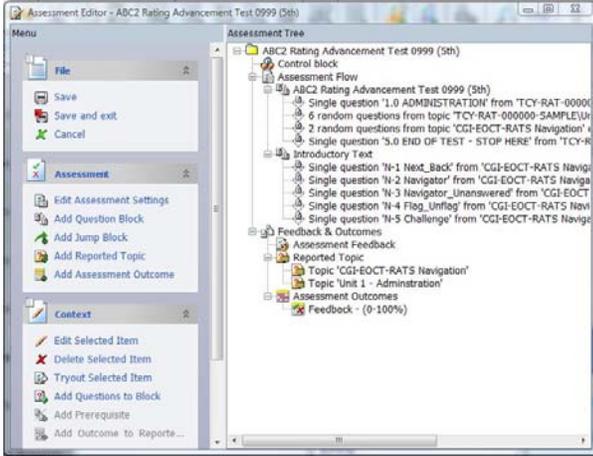
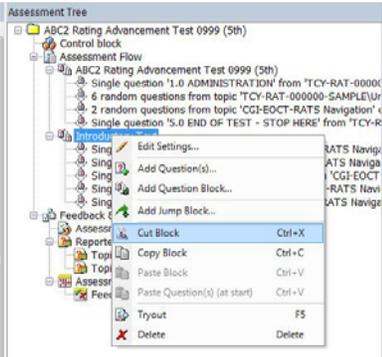
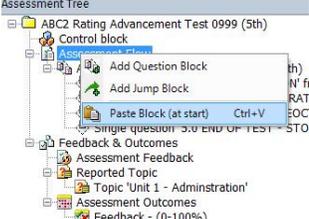
Step	Action
13	<p>Type <i>Introductory Text</i> into the Block Name.</p> <p>✓ Check <i>Disable feedback in this block</i></p> <p>✓ Check <i>Exclude this block from the assessment time limit</i></p> <p>✓ Check <i>Use different template in this block</i></p> <p>Select USCG_EOCT_RAT_INTRO from the Alternative template file name for this block dropdown list.</p> 
14	<p>Left Click <i>Introduction</i>. The <i>Question Block Settings</i> dialog screen will open. This is where you will add the rating badge corresponding to the RAT and CGI introductory text.</p>
15	<p>Position your cursor inside of the dialog box  and left click the add picture icon:</p> 

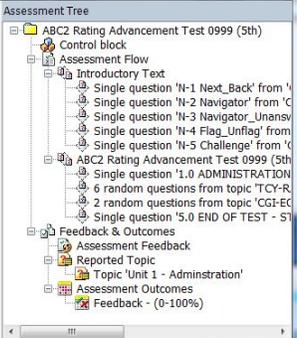
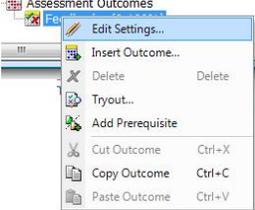
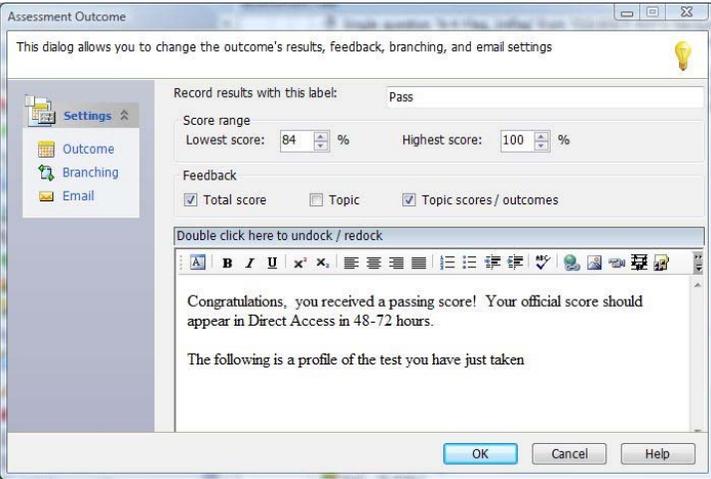
Step	Action
16	<p>Click <i>Browse...</i></p> <p>The Insert Picture location dialog box will open. You will see your list of topics.</p> <p>Select <i>Global Resources</i>.</p> <p>A list of Global Resources will display.</p> 
17	<p>Double Left Click <i>uscg_rating_badges</i>.</p> <p>The list of rating badges available will open.</p>  <p>Scroll down until the desire rating badge is located.</p>
18	Highlight the desired badge and click <i>Open</i> .
19	The file location of the rating badge will populate the dialog box.

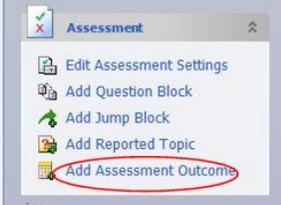
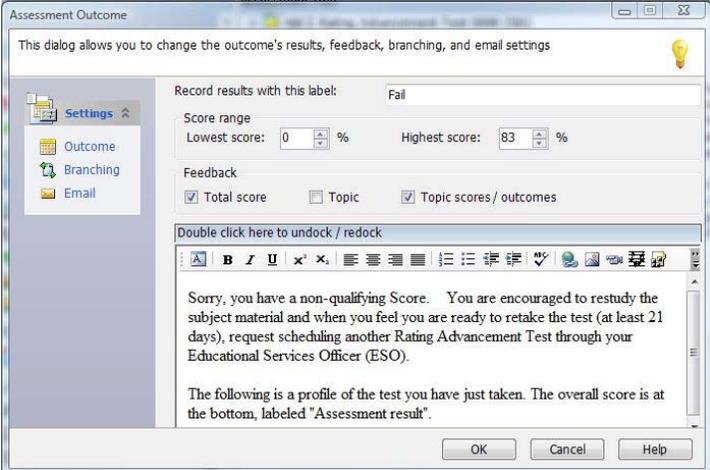
Step	Action
20	<p>Type Rating Badge in the Alternate Text box.</p> <p>Under Layout, set the Alignment to Left.</p> <p>Leave Border Thickness set to zero (0)</p> <p>Under Spacing – set both Horizontal and Vertical to “4”</p> 
21	Click OK . The rating badge will be set in the upper left of the introduction screen.
22	Copy the text shown in step 24 by highlighting and pressing <Ctrl>C.
23	<p>Position the cursor next to the rating badge and press <Ctrl>V to paste the text.</p>  <p>Note: Be sure to proofread the text and change the test time limit to match your test.</p>

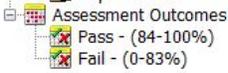
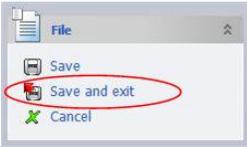
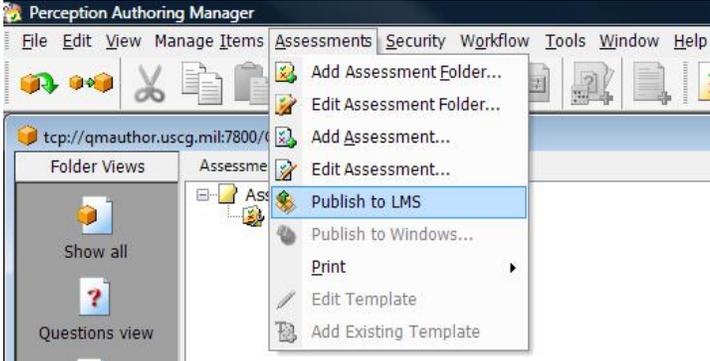
Step	Action
24	<p>Enter the following text into the Introduction block:</p> <p>PARTICIPANT INSTRUCTIONS</p> <ol style="list-style-type: none"> 1. You are not permitted to bring any personal items into the examination room: this includes all bags, notes, cell phones, pagers, cameras, pens, and books (unless authorized by the CG Institute). 2. Listed after each item in this test are four answers. Only one of the answers is correct. 3. When you have decided on the correct answer to the test item, select the appropriate answer by clicking on it with your mouse. 5. Challenges to Test Items: If you feel an item is incorrect and desire to challenge it, click on the CHALLENGE button and enter the required information. The CHALLENGE button is located in the bottom of the navigator box. <p>ANSWERS FOR THE ITEMS IN THIS TEST ARE BASED ON THE CURRENT RPQ's</p> <ol style="list-style-type: none"> 6. Scratch paper is accountable and shall be returned to the test administrator upon completion of your exam. 7. There is a XXX MINUTE TIME LIMIT for this test. Once you have started you must continue answering the items until the test is complete. 8. This Test is for OFFICIAL USE ONLY. You are not to discuss the content of this test with any other person after leaving the test site. Disclosing any contents of this test, memorizing and later transcribing from it, receiving assistance of any kind while taking the test or any combination of these is prohibited and punishable under the Uniform Code of Military Justice (UCMJ). <p>Click -NEXT- now to receive a brief overview on how to navigate through this exam, or -SUBMIT- to begin the exam. The timer will not start until you click -SUBMIT-</p> <p>By selecting submit and proceeding to the test I acknowledge that I have read and understand that I am not to discuss the content of this test with any other person after leaving the test site. Disclosing any contents of this test, memorizing and later transcribing from it, receiving assistance of any kind while taking the test or any combination of these is prohibited and punishable under the Uniform Code of Military Justice (UCMJ).</p>

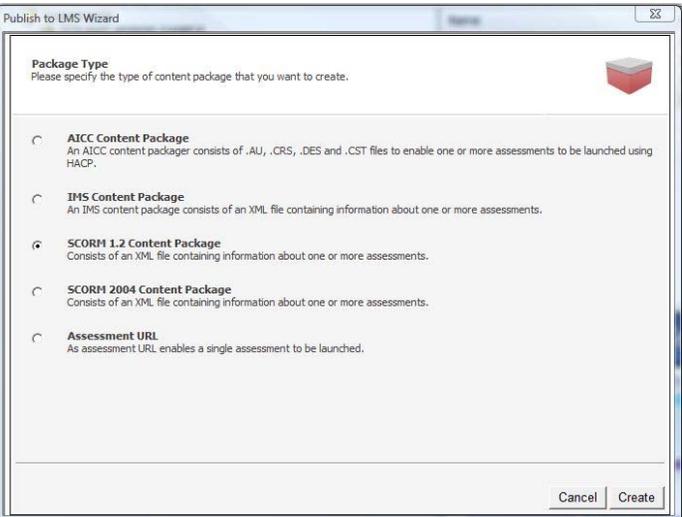
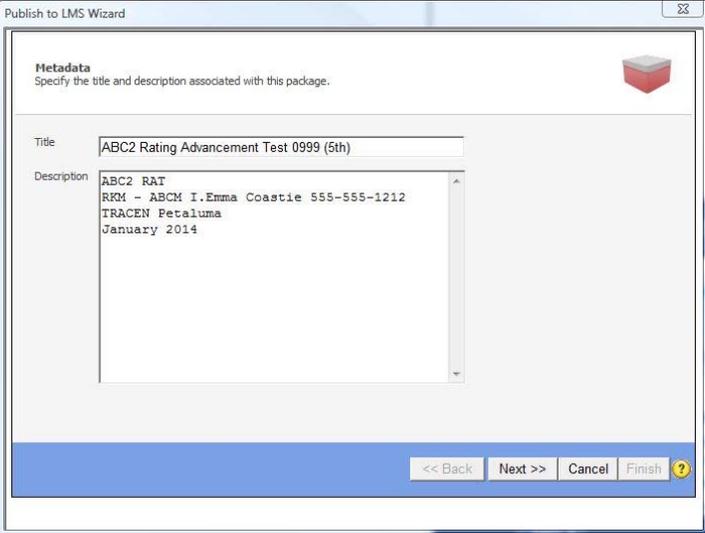
Step	Action
25	<p>Click OK. The Select Questions dialog box will open.</p> 
26	<p>If your local testing system administrator has set your permissions correctly you should have “View/Use” permission to the CGI-EOCT-RATS Navigation topic folder which should now be visible in the Select Questions dialog box.</p> <p>Expand the CGI-EOCT-RATS topic by clicking on “+” next to five items should appear.</p>  <p style="text-align: right;"><i>Navigation</i> the title –</p>
27	<p>Click on Single question in the menu box.</p>
28	<p>Double click on the CGI Navigation items, (N-1, N-2, N-3, N-4, N-5), in order, time until are added to dialog box.</p>  <p style="text-align: right;">one at a all five the</p>
29	<p>Click OK.</p>

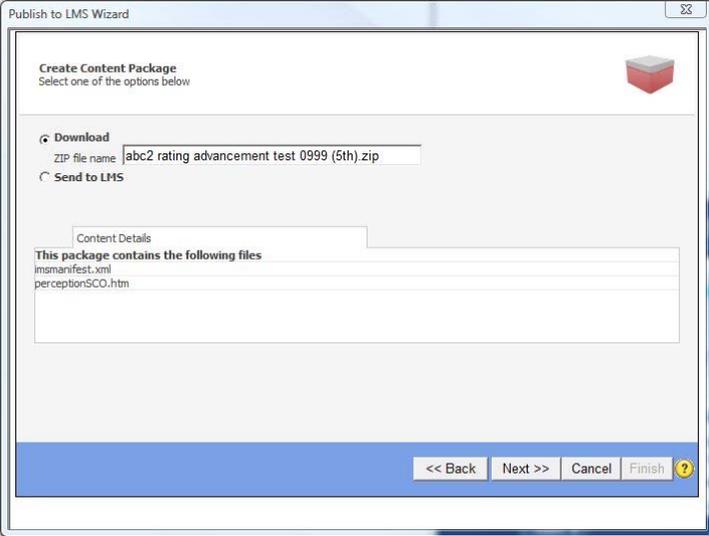
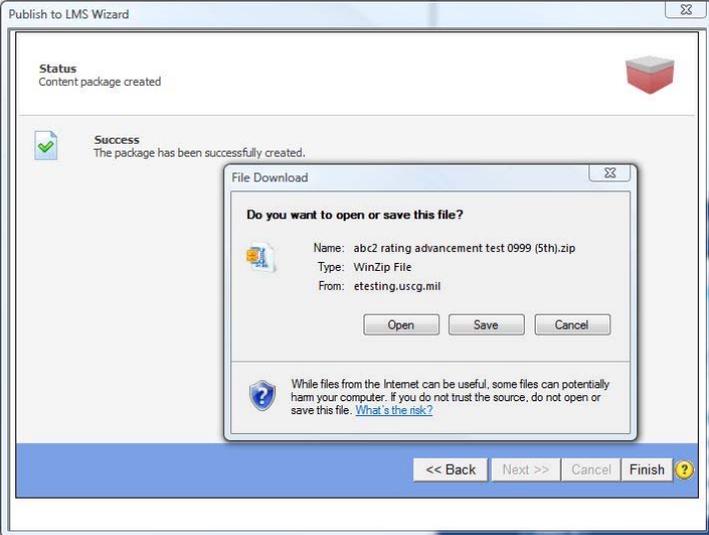
Step	Action
30	<p>The Assessment Tree dialog box will now appear.</p>  <p>You will notice the new “Question Block” has been added after the actual RAT. The new Question Block must now be moved to the beginning of the test.</p>
31	Click on Introductory Text to highlight it.
32	<p>Right click. A dropdown menu will open. Select Cut Block.</p> 
33	Left click on Cut Block.
34	The Assessment Introduction block will disappear!
35	Move the cursor to Assessment Flow just above the assessment name. Highlight the Assessment Flow. Right click.
36	<p>A dropdown box will open. Paste Block (at start).</p> <p>Left click.</p>  <p>Select</p>

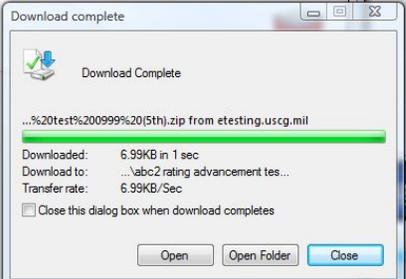
Step	Action
37	<p>The introduction block will be placed before the assessment the Assessment Flow.</p> <p>Next we will set the cut score.</p>  <p>now be block in (passing)</p>
38	<p>Highlight Feedback under Outcomes on the Assessment click to reveal a menu box. Settings and left click.</p>  <p>Assessment Tree, right Select Edit</p>
39	<p>The Assessment Outcome dialog box for a passing score will open. Ensure “Record results with this label:” reads Pass</p> <p>Set the Score range to Lowest score = the cut score of the RAT</p> <p>Set the Highest score to 100%</p>
40	<p>Set Feedback to</p> <p><input checked="" type="checkbox"/> Total score <input type="checkbox"/> Topic <input checked="" type="checkbox"/> Topic scores/outcomes</p>
41	<p>Enter the following congratulatory message in the text block :</p> <p>Congratulations, you received a passing score! Your official score should appear in Direct Access in 48 – 72 hours.</p> <p>The following is a profile of the test you have just taken:</p> 
42	<p>Click OK</p>

Step	Action
43	You will now set the failing score and message.
44	Highlight <i>Assessment Outcomes</i>
45	Left Click on <i>Add Outcome</i> from the menu.  <i>Assessment</i>
46	The Assessment Outcome dialog box for a failing score will open. Ensure “ <i>Record results with this label:</i> ” reads Fail Set the Score range to Lowest score = 0% Set the Highest score to one point below RAT cut score
47	Set Feedback to <input checked="" type="checkbox"/> Total score <input type="checkbox"/> Topic <input checked="" type="checkbox"/> Topic scores/outcomes
48	Enter the following in the text box: Sorry, you have a non-qualifying score. You are encouraged to restudy the subject material and when you feel you are ready to retake the test (at least 21 days), request scheduling another Rating Advancement Test through your Educational Services Officer (ESO). The following is a profile of the test you have just taken. The overall score is at the bottom, labeled "Assessment result." 
49	Click OK

Step	Action
50	<p>You have almost completed the process! You should not see two Assessment Outcomes, Pass and Fail at the bottom of the Assessment Tree:</p> 
51	<p>This is the most important click Save and Exit in the menu!</p>  <p>step...Left</p>
52	<p>The online version is now ready to be published to the Learning Management System.</p> <p>Highlight the assessment and left click on <u>A</u>ssessments on the top menu bar.</p> <p>A dropdown box will open.</p>  <p>Select <i>Publish to LMS</i></p>
53	<p>The <i>Publish to LMS Wizard</i> will open.</p>

Step	Action
54	<p>Select <i>SCORM 1.2 Content Package</i></p> <p>Click <i>Create</i></p> 
55	<p>The Metadata dialog box will open.\</p> <p>Enter the title of your RAT exactly the same as you titled it when you created it.</p> <p>Enter a brief description which includes the short title, RKM/ARKM contact info and effective date.</p> 
56	Click <i>Next</i>
57	<p>The second Assessment Metadata screen will open.</p> <p>Click <i>Next</i></p>
58	<p>The Package Keywords dialog box will open</p> <p>Click <i>Next</i></p>

Step	Action
59	<p>The Create Content Package dialog box will open.</p> <p>Check to be sure the Zip file name is populated and the Content Details list contains two files:</p> <p>imsmanifest.xml</p> <p>perceptionSCO.htm</p>  <p>If all appears OK, click Next</p>
60	<p>The File Download dialog box will open. Select “Save”.</p> 
61	<p>The Save As window will open. Select a save location just as you would for any file and click Save.</p>

Step	Action	
62	<p>The Download dialog box will open the save status. indicates “Download Complete”, click</p>	 <p>Complete and show When it</p> <p>Close</p>
63	Click Finish in the remaining Publish to LMS Wizard window	
64	Contact the LMS manager at FC-Tadl for instruction to upload and activate the RAT on the LMS.	
65	<p>You have now completed the process of developing and preparing your RAT for online administration. Contact the Non-resident training branch at the Coast Guard Institute for additional assistance or guidance.</p>	

THIS PAGE LEFT INTENTIONALLY BLANK

Appendix C

JOB AID FOR THE MODIFIED ANGOFF METHOD

Applying the Modified Angoff Method

Background

The main rationale behind criterion-referenced cutoff scores is that one must be able to distinguish between candidates who can demonstrate sufficient knowledge to be advanced, licensed, or certified and those who cannot. The cutoff score creates two classifications of candidates: those who are competent to perform the duties in a safe and effective/definitive manner and those who may not perform at this level.

The Angoff method defines the cutoff score as the lowest score the minimally acceptable candidate is likely to achieve. Candidates scoring below this level are believed to lack sufficient knowledge, skills, or abilities to be certified. (Moritsch, 2000)

Step #1: Select/Gather the Raters

Select five or more Subject Matter Expert (SME) raters. Ideally, the number of raters should be 8 - 10 to keep the variance among ratings low; the more raters involved, the more accurate the cut score will be.

Note: All raters are required to execute a non-disclosure statement that will become part of their official personnel record to discourage possible compromise of assessment items. The standard non-disclosure statement is included as Appendix E in this SOP.

A SME is a person most knowledgeable regarding a specific subject, but not necessarily the person with the most practical experience or best performance (i.e., a person who has knowledge of and has performed a wide variety of tasks pertaining to the subject area but is not performing those tasks at this time).

When selecting SME raters, identify those who may have received awards or recognition for their performances, and solicit recommendations from peers and supervisors. Raters should possess the following proficiencies:

- Familiarization with the tasks the test will assess
 - Knowledge of the skill sets of persons who will perform those tasks
 - Ability to pass the existing test at the current cut score
 - Ability to edit test items for clarity, accuracy, spelling, and grammar
-

Applying the Modified Angoff Method (continued)

Step #1: Select/Gather the Raters (continued)

Additionally...

- For Rating Advancement Tests (RAT), choose SMEs who are at least one grade higher than the examinees (i.e., for E-5 and E-6 RATs, use E-7, E-8, and/or E-9 SMEs).
- For “C” school assessments, merchant mariners’ examinations, flight crew, etc., the judges should be selected from a pool of successful course graduates, instructors, licensed merchant mariners, or flight crew who are considered experts on the subjects in the assessment.
- Try to assemble a diverse group of SMEs (e.g., different races, genders, ages, educational backgrounds, geographical locations, etc.).

Gather the raters at a common location where they can work both independently and together. Have the raters take the test using the current cut score, if one has been established, using same conditions of testing as “real” test takes would. This establishes the “ceiling” for the raters item score. The “ceiling”, if established is the highest rating each item may be assigned because if the “experts” can only achieve this average score, we assume the minimally qualified performer will not be able to achieve this score. Obtain feedback from raters on wording, and design of test items. If items need to be revised, do so before the rating process begins.

Note: In the case of a large test item data bank it may not be practical for the raters to complete the entire item bank due to time constraints so this step may be omitted and noted in the test plan.

Applying the Modified Angoff Method (continued)

Step #2: Identify Minimally Acceptable Candidate (MAC)

Prior to rating test items, the SME raters must agree on what constitutes a minimally qualified performer. If agreement cannot be reached, do not attempt to apply Angoff because the test item rating results will be too varied, resulting in a distorted cut score.

In order to distinguish between those who are competent to do the job and those who are not, we must first define levels of competency as they relate to observable work behaviors. There are two levels of competency that must be defined.

A *Minimally Acceptable Candidate (MAC)* describes what the minimally acceptable candidate should be able to do, or should know, on the very first day on the job. The minimally acceptable candidate should be defined according to observable work behaviors through which a person can demonstrate his/her knowledge, skills, or abilities.

A *Superior Worker* describes what the superior, or highly qualified person should be able to do, or should know, on the very first day on the job. The superior worker should be defined according to observable work behaviors through which a person can demonstrate his/her knowledge, skills, or abilities.

Since the Modified Angoff Method is based on the ability of a “minimally acceptable candidate” to answer test items, it is necessary to identify competencies that represent such a person prior to rating the test items.

A minimally qualified performer is:

- One who performs the task in the field; **not** a student
- One who has the least amount of education and experience necessary to perform the task
- One who meets standards, though barely
- One whose task performance is borderline, but acceptable

In addition to the criteria listed above, factors specific to the job/tasks may be introduced to further identify a minimally qualified performer.

For example, a **minimally qualified** boarding officer might be:

- One who knows how to communicate with and interview citizens
 - One who knows how to use the Boarding Officer Job Aid Kit
 - One who knows how to complete required forms
-

Applying the Modified Angoff Method (continued)

**Step #2:
Identify
Minimally
Acceptable
Candidate
(MAC)
(continued)**

Another CG Rating related example that ranks a performer from not competent to highly competent may be:

- Machinery Technician
 - Job: Replace the carburetor on a small engine
 - Not competent – “What’s a carburetor?”
 - **Minimally competent** – “OK – but I may need some help. Can you stick around?”
 - Competent – “OK – But if I get stuck can I call you?”
 - Superior (highly competent) – “No problem, I’ll let you know when I’m done.”
-

Applying the Modified Angoff Method (continued)

Step #3: Rate the Items

Test item rating should be done independently by the raters. Allow approximately two hours for a 100-item test. To rate the test items, proceed as follows:

Step	Action
1	<p>Each judge/rater is given a copy of all test items. Each judge/rater should complete the assessment as a test taker would as this will:</p> <ul style="list-style-type: none"> • Verify validity of the items • Establish a “ceiling” cut score <p>When rating an entire database used to draw random test items it may not be practical for raters to “take” the exam so that may be skipped and documented.</p> <p>Do NOT provide raters with the answer key. This could unduly influence the raters by causing them to underestimate the item difficulty.</p>
2	<p>Have raters review each test item, one at a time, and estimate the number of minimally competent candidates out of 100 who would answer the question correctly. They can record their “scores” for each item directly on the test or on a separate form or spreadsheet provided by the facilitator.</p> <p>The raters should examine each stem and all distractors and decide how many distractors a minimally competent candidate would surely eliminate. This will establish a “floor” to rating the item:</p> <ul style="list-style-type: none"> • If one distractor on a 4-alternative item will most likely be eliminated, then the lowest possible rating for the item would be .33 because there are now only 3 logical choices remaining. • If the stem and all distractors have been properly constructed, and all of the distractors should be plausible so the floor would be .25. <p>Note: Raters should periodically refer to the competencies of the minimally qualified performer to ensure their rating estimates are as accurate as possible.</p>
3	<p>Record the rating estimate on the Test Item Rating form under the “Percentage Correct” (i.e., if 80 out of 100 minimally qualified performers would answer a question correctly, then the percentage correct is 80%).</p> <p>Note: Estimates should not be higher than 95 nor lower than 25 for a 4-alternative item (33 for a 3-alternative item). Not even strong performers are expected to earn a perfect score of 100; and minimally qualified performers can correctly guess an answer 25% of the time for a test item with 4-alternatives (33% for a 3-alternative item)</p> <p>Theoretically, any value between 25 and 95 is acceptable but for simplicity it is recommended to use a set scale such as: 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95.</p>

Applying the Modified Angoff Method (continued)

Step #4: Review the Ratings

Collect the raters’ Test Item Rating forms and enter the results in the Ratings Results spreadsheet (figure C-1) which is available, with instructions on the CG Portal Sharepoint site.

- Enter the percentages for each test item under the respective rater’s name. As each rating is entered the spreadsheet will calculate the difficulty, standard deviation and Angoff rating for each item automatically.

	Test Item QID	Enter * If New, R If Retired	Difficulty Metatag	Average Percentage Correct (Angoff Rating)	Expert 1 Name	Expert 2 Name	Expert 3 Name	Expert 4 Name	Expert 5 Name	Expert 6 Name	Expert 7 Name	Expert 8 Name	Standard Deviation	Unit Cut Score	70.00	Moderate Difficulty
7																
8	6.2.1.171		Moderate	59.29	65	60	60	65	55	65	45		7.32			
9	6.2.1.172		Moderate	72.86	75	70	85	75	55	75	75		9.06	25 - 49.99	Hard	
10	6.2.1.173		Moderate	65.71	65	65	60	80	55	75	60		8.86	50 - 74.99	Moderate	
11	6.2.1.201		Moderate	50.71	50	50	50	65	45	45	50		6.73	75 - 95	Easy	
12	6.2.1.202		Moderate	60.00	65	60	60	75	55	55	50		8.16			
13	6.2.1.203		Moderate	64.29	65	70	65	80	55	55	60		8.86			
14	6.2.2.171			81.43	80	75	80	90	85	75	85		5.56			
15	6.2.2.172		Moderate	57.86	50	55	65	65	65	55	50		6.99			
16	6.2.2.301		Moderate	66.43	75	75	60	75	65	55	60		8.52			
17	6.2.2.302		Moderate	57.86	65	55	45	65	55	65	55		7.56			
18	6.2.2.541		Moderate	72.86	75	70	75	90	65	75	60		9.51			
19	6.2.2.502			77.14	65	70	80	85	75	75	90		8.59			
20	6.2.2.503		Moderate	74.29	75	65	75	75	65	75	90		8.38			
21	6.2.2.544			78.57	75	80	75	80	85	75	80		3.78			
22	6.2.2.505			79.29	75	80	75	75	95	75	80		7.52			
23	6.2.3.111			77.14	75	65	70	85	95	75	75		9.94			
24	6.2.3.102			78.57	75	70	75	95	75	75	85		8.52			
25	6.2.3.201			77.86	80	65	80	80	85	65	90		9.51			

Difficulty Rating

25 - 49.99 Hard

50 - 74.99 Moderate

75 - 95 Easy

Standard Deviation

A standard deviation of more than 10 will trigger an alert. Discuss the outliers with the judges who set them to determine why. Change as necessary.

19 Easy in this unit 20%

54 Moderate in this unit 73%

1 Hard in this unit 1%

74 TOTAL 100%

Figure C-1. Expert Ratings Spreadsheet Example

Different estimates from raters for the same test item are to be expected. Arbitrariness can result from diverse conceptions of mastery of the task, various interpretations of the learning objectives, misunderstanding of the test item etc.

Standard deviation reflects the amount of agreement/disagreement among the raters for each test item. A low standard deviation indicates a high agreement among raters. A high standard deviation is grounds for further examination of that test item.

Convene the Raters

For any test item whose standard deviation exceeds 10, raters should discuss the reasons for variations in the estimates. The intent of the discussion is to increase agreement among the raters. By discussing how the raters arrived at such different conclusions for a test item, they might decide to re-evaluate their estimates.

Re-Evaluate Test Items

After discussion, separate the raters once again and have them rate any test items with standard deviations above 10. Collect the ratings and enter them on the spreadsheet. (This step can be completed with open discussion using caution that raters are not unduly influenced by dominate members.)

Applying the Modified Angoff Method (continued)

Step #4: Review the Ratings (continued)

Once test items have been re-evaluated and the estimates have been re-entered into the Ratings Results spreadsheet, review the sheet for the following:

- If a rater provided the same rating for every test item, consider eliminating those ratings.
- If a rater continually provided ratings that were very dissimilar from the other raters, consider eliminating those ratings.
- If an outlying standard deviation for a test item remains, consider another discussion/re-evaluation session.

Note: Even if disagreement persists, the average percentage for that test item can be factored into the cut score.

Step #5: Determine the Final Cut Score

As the rating for each test item in each unit/topic are entered into each page (worksheet), the spreadsheet will compile the the numbers and generate the final test design blueprint (figure C-2) to design a fair, randomized test.

Final Randomized Test Design Blueprint for: SAMPLE															25-Sep-14		
Unit	Unit Cut Score & Difficulty	Items in Unit	% of Total Items	Avail Hard	% From Unit	Avail Mod	% From Unit	Avail Easy	% From Unit	Total # Needed From Unit	Use Hard (Calculated)	Use Hard (Actual)	Use Mod (Calculated)	Use Mod (Actual)	Use Easy (Calculated)	Use Easy (Actual)	Unit
1.0	75	9	6.67%	0	0%	5	56%	4	44%	3.33	0.00	0	1.85	2	1.48	1	1.0
2.0	70	74	54.81%	1	1%	54	73%	19	26%	27.41	0.37	1	20.00	20	7.04	7	2.0
3.0		0	0.00%	0		0		0		0.00		0		0		0	3.0
4.0	70	43	31.85%	0	0%	33	77%	10	23%	15.93	0.00	0	12.22	12	1.70	3	4.0
5.0		0	0.00%	0		0		0		0.00		0		0		0	5.0
6.0	65	9	6.67%	1	11%	6	67%	2	22%	3.33	0.37	1	2.22	2	0.74	1	6.0
7.0		0	0.00%	0		0		0		0.00		0		0		0	7.0
8.0		0	0.00%	0		0		0		0.00		0		0		0	8.0
9.0		0	0.00%	0		0		0		0.00		0		0		0	9.0
TOTAL		135	100.00%	2		98		35		50.00	0.74	2	36.30	36	12.96	12	

NOTE: If appears in the "Total # Needed From Unit" block - you do not have sufficient items in the unit indicated to design a fair test.

Test Difficulty Moderate

Test Cut Score

70.00

Set Desired Test Size

50

Checksum

50

After all Angoff data has been entered on Unit worksheets set the desired test size in the block to the left. Based upon the number of available items, the quantity of Hard, Moderate and Easy from each unit will populate automatically. Use these results to design the test in QMP using established SWE/RAT Metatags .

Note: Due to rounding errors in Excel, the unit/item difficulty totals may require you to round up or down manually to achieve desired test size. Set the actual number desired based upon the calculated results in the columns labeled "Actual" above. The Checksum to the left will alert you if the selected value does not match the desired test size.

Figure C-2. Calculating the Cut Score

Note: If the calculated cut score is higher than the average score obtained by the raters when they completed the test (if that step was not omitted), the cut score must be reduced as a minimally competent candidate cannot be expected to achieve a score higher than an “expert.”

Notes Concerning the Modified Angoff Method

Test Item Difficulty Rating

The Ratings Results spreadsheet includes a column for identifying the difficulty rating for each test item. Test items can be rated as Hard, Moderate, or Easy.

This will be automatically revealed by the overall rating score for each test item according to the scale below (Figure C-3).

Difficulty Rating	
25 - 49.99	Hard
50 - 74.99	Moderate
75 - 95	Easy

Figure C-3. Difficulty Rating

Fairness in Randomization

When working with a large test item database which selects items from several topics on a random basis, it is important to generate each test instrument as fairly as possible. In addition to selecting the same number of items from each topic (EO), items of similar difficulty must be selected. To accomplish this it is necessary to tag each item in the test item database with its difficulty established during the Angoff rating. See figure C-4.

RAT/SWE Difficulty	Description	Order	Status	Date modified	Edi...	Workfl...
Moderate (50 - 74)	QID 6.2.1.3/1 What cl...	7	Normal	Jan 08 2013 14:...	Ja...	(USCG...
Moderate (50 - 74)	QID 6.2.1.3/10 Which...	4	Normal	Jan 17 2013 12:...	Ma...	(USCG...
Moderate (50 - 74)	QID 6.2.1.3/2 Plywoo...	2	Normal	Jan 08 2013 14:...	Ja...	(USCG...
Moderate (50 - 74)	QID 6.2.1.3/3 Lumber...	8	Normal	Jan 03 2013 09:...	Ja...	(USCG...
Easy (75 - 95)	QID 6.2.1.3/4 What le...	1	Normal	Jan 24 2013 13:...	Ma...	(USCG...
Easy (75 - 95)	QID 6.2.1.3/5 What ty...	10	Normal	Dec 21 2012 13:...	Ma...	(USCG...
Moderate (50 - 74)	QID 6.2.1.3/6 How sh...	5	Normal	Jan 08 2013 14:...	Ja...	(USCG...
Moderate (50 - 74)	QID 6.2.1.3/7 Of the f...	9	Normal	Jan 02 2013 12:...	Ma...	(USCG...
Easy (75 - 95)	QID 6.2.1.3/8 What pi...	3	Normal	Jan 17 2013 12:...	Ma...	(USCG...
Easy (75 - 95)	QID 6.2.1.3/9 The siz...	6	Normal	Jan 17 2013 12:...	Ma...	(USCG...

Figure C-4. Difficulty Tags

Once this is accomplished, use of the spreadsheet to calculate the number of test items in each section, by difficulty and percentage of test size should be used to design the final test. See figure C-5 for an example.

Total # Needed From Unit	Use Hard (Calculated)	Use Hard (Actual)	Use Mod (Calculated)	Use Mod (Actual)	Use Easy (Calculated)	Use Easy (Actual)
3.33	0.00	0	1.85	2	1.48	1
27.41	0.37	1	20.00	20	7.04	7
0.00		0		0		0
15.93	0.00	0	12.22	12	3.70	3
0.00		0		0		0
3.33	0.37	1	2.22	2	0.74	1
0.00		0		0		0
0.00		0		0		0
0.00		0		0		0
50.00	0.74	2	36.30	36	12.96	12

Figure C-5 – Final Test Design

Appendix D

TEST PLAN

Sample Test Plan

TESTING SENSITIVE – HANDLE AS FOUO
Barista Second Class (BA2)
Rating Advancement Test Plan

QPM Exam Location: (Repository Address and file name go here)
Exam Title: BA2 Rating Advancement Test 0216 (1st)
Effective Date: 01 August 2013
Exam Type: Criterion-Referenced
Exam Level: High Stakes
Cut Score: 75%
Number of Test Items: 50
Time Limit: 100 Minutes
Rate Knowledge Manager (RKM): BACM I. M. Coastie (555)-555-1212
(Ida.M.Coastie@uscg.mil)

Statement of Purpose:

1. The Barista Second Class (BA2) Rating Advancement Test (RAT) is a high stakes, criterion-referenced, examination designed to test the knowledge and application of each of the BA2 Enlisted Rating Performance Qualifications (RPQ).
2. The construct of the test is 50 item, 4-choice multiple choice. Only one response is considered correct. A mechanism for the test taker to challenge the validity of a test item is provided. A hyperlink to selected references available for the test taker to research will be provided during the test. If the test taker chooses to use the references the 100 minute time limit does not pause. A Compact Disk - Read Only Memory (CD-ROM) with references will be provided to all Educational Services Officers (ESOs) as a backup.
3. The BA2 RAT is designed to be delivered electronically, via computer, managed and scheduled within the Coast Guard Learning Management System (LMS). Three paper-based exams backup will be available as a contingency and will be administered as per current Coast Guard Institute (CGI) instructions for the administration of End of Course Tests (EOCTs) and Rating Advancement Tests (RATs).
4. A satisfactory score is required to be considered at a minimally acceptable competency (MAC) level to be eligible for advancement from BA3 to BA2 and is a prerequisite for consideration for advancement.
5. The US Coast Guard Human Resources Directorate (CG-1), established the requirement for the Rating Advancement Test as a prerequisite to advancement, unless waived by proper authority.

TESTING SENSITIVE – HANDLE AS FOUO
Barista Second Class (BA2)
Rating Advancement Test Plan

Analysis of Job Content:

1. This examination is designed to support the required job functions of a Barista Second Class (BA2) in four specific areas of duty: Administration, Food Safety, Operations, and Customer Relations. Specific topics within each of these areas were identified as those performed by a majority of the BA2 paygrade by an Occupational Analysis (OA) final report dated 09FEB2012 performed by the USCG Performance Technology Center. A copy of this analysis is available on the Coast Guard Intranet at:
<http://cgweb.tcyorktown.uscg.mil/PTC/Library/SSIC/1000/XXXXXX//2012BAFinalReport.pdf>

Content Validity of Objectives:

1. The content of the exam questions was derived from the tasks identified as those required to be performed by a BA2.
2. References and guidance for accomplishment of the tasks required are published in the Rating Performance Qualification Standard (RPQS) for Barista Second Class Petty Officer, form CGRPQS-BA2 (MAY/2014) which is available via the Coast Guard Learning Management System at:
<https://elearning.uscg.mil/catalog/course.asp?id=XXXXXXX>.
3. All areas of the OA were reviewed by a panel of nine accomplished performers (APs) selected from various operational units throughout the Coast Guard. The APs were selected by their respective commanding officers because they have demonstrated superior knowledge and performance of the job requirements of the BA rating. The APs consisted of paygrades, E-5 through E-7. Names are not published due to privacy concerns.

Cognitive Items/Rating Instruments:

1. The RAT is supported by the following Performance Objectives: (Objectives have been abbreviated to conserve space)

Unit 5.1.0 –Administration

Task 5.1.1 – PREPARE XXXXXXXXX

- 5.1.1.1 – DRAFT XXXXXXXX
- 5.1.1.2 – ESTIMATE XXXXXXXX
- 5.1.1.3 – ESTIMATE XXXXXXXX
- 5.1.1.4 – ESTIMATE XXXXXXXX
- 5.1.1.5 – ESTIMATE XXXXXXXXt
- 5.1.1.6 – SUBMIT XXXXXXXXX

Task 5.1.2 – MANAGE XXXXXXXXXXXXXXXX

Page 2 of 8

TESTING SENSITIVE – HANDLE AS FOUO

TESTING SENSITIVE – HANDLE AS FOUO
Barista Second Class (BA2)
Rating Advancement Test Plan

- 5.1.2.1 – REVIEW xxxxxxxxxxxx
- 5.1.2.2 – CREATE xxxxxxxxxxxx
- 5.1.2.3 – TRACK xxxxxxxxxxxx
- 5.1.2.4 – GENERATE xxxxxxxxxxxx

Unit 5.2.0 – Food Safety

Task 5.2.1 – MAINTAIN xxxxxxxxxxxx

- 5.2.1.1 – IDENTIFY xxxxxxxxxxxx
- 5.2.1.2 – VERIFY xxxxxxxxxxxx
- 5.2.1.3 – NOTIFY xxxxxxxxxxxx
- 5.2.1.4 – INSPECT xxxxxxxxxxxx
- 5.2.1.5 – TEST xxxxxxxxxxxx
- 5.2.1.6 – COMPLTE xxxxxxxxxxxx

Task 5.2.2 – MANAGE xxxxxxxxxxxx

- 5.2.2.1 – VALIDATE xxxxxxxxxxxx
- 5.2.2.2 – VALIDATE xxxxxxxxxxxx
- 5.2.2.3 – CONDUCT xxxxxxxxxxxx
- 5.2.2.4 – COMPLETE xxxxxxxxxxxx
- 5.2.2.5 – CONDUCT xxxxxxxxxxxx

Unit 5.3 – Operations

Task 5.3.1 – DESIGN xxxxxxxxxxxx

- 5.3.1.1 – DETERMINE xxxxxxxxx
- 5.3.1.2 – DETERMINE xxxxxxxxx
- 5.3.1.3 – ESTIMATE xxxxxxxxx
- 5.3.1.4 – SELECT xxxxxxxxx

Unit 5.4 – Customer relations

Task 5.4.1 – INTERPRET xxxxxxxxxxxx

- 5.4.1.1 – LOCATE xxxxxxxxx
- 5.4.1.2 – IDENTIFY xxxxxxxxx
- 5.4.1.3 – TRANSLATE xxxxxxxxx
- 5.4.1.4 – TRANSLATE xxxxxxxxx

Task 5.4.2 PERFORM xxxxxxxxxxxx

TESTING SENSITIVE – HANDLE AS FOUO
Barista Second Class (BA2)
Rating Advancement Test Plan

- 5.4.2.1 – PREPARE xxxxxxxxxxxxxxxx
 5.4.2.2 – SET UP xxxxxxxxxxxxxxxx
 5.4.2.3 – DISPLAY xxxxxxxxxxxxxxxx
 5.4.2.4 – EJECT xxxxxxxxxxxxxxxx
 5.4.2.5 – STOW xxxxxxxxxxxxxxxx

2. The RAT is designed to support the Rating Performance Qualification Standard (RPQS) for Barista Second Class Petty Officer, form CGRPQS-BA2 (MAY/2014).
3. The test items were designed by the BA Rate Knowledge Manager (RKM) assigned to USCG Training Center Anytown, USA.
4. The RKM is a Master Chief Petty Officer (MCPO) at a military pay grade of E-9 and is considered a subject matter expert and rating course development specialist by virtue of his or her superior rating knowledge, experience and assignment.
5. The test items for this examination were developed and written during the period 01 January 2013 through 05 July 2013.

Content Validity of Items and Test Instruments:

1. All test items and the RAT are designed to support Rating Performance Qualifications (RPQ) for BA2s listed on form CGRPQS-BA2 (MAY/2014). All test items were entered into a workflow process within the Coast Guard enterprise e-testing authoring, delivery and analysis tool, Questionmark Perception (QMP). The workflow is illustrated below:



The test item with distracters is created by the RKM (Authoring). The author forwards the test item to a writer/editor (W/E) who checks the item for correct grammatical construction, parallelism of correct response and distracters and validity of references. The item is forwarded to an Instructional Systems Specialist (ISS) for final review and either approved for use or rejected for revision.

2. The test items were validated using the workflow process during the period 01 January 2013 through 05 July 2013.
3. Records of the validation process are maintained by the QMP workflow history function within the application.

TESTING SENSITIVE – HANDLE AS FOUO
Barista Second Class (BA2)
Rating Advancement Test Plan

Initial Test Pilot:

1. A pilot of test functionality was conducted on 05 August 2013.
2. The pilot was conducted at USCG Training Center Anytown, USA.
3. The pilot was conducted by the Test Development Manager, Test Writer/Editor and the RKM who read each test item and all responses, tested any links embedded in the test item and made any notes regarding suggestions or inoperability of functions.

Item Analysis:

1. There was no item analysis performed during the pilot.
2. As data is gathered each test item will be evaluated for validity, reliability, difficulty and discrimination. Items that do not perform within industry accepted standards will be retired.
3. Intentions are to insert non-counting test items into the production examinations at regular intervals to gather item analysis data for future use.

Parallel Forms and Item Banks:

1. The intent is to administer this assessment via computer in most cases. Each form of the test will be designed randomly with the same number of items with the same level of difficulty from each unit of instruction selected for each form of the examination. Test items will be distributed randomly within the test block. Paper examinations will be produced in the same manner and sent to testing units as a backup should computer access become unavailable.
2. The test will be 50 question, multiple choice format designed as follows:
 - Unit 1.0: 29% - 15 items
 - Hard: 0
 - Moderate: 5
 - Easy: 10
 - Unit 2.0: 21% - 10 items
 - Hard: 0
 - Moderate: 5
 - Easy: 5
 - Unit 3.0: 18% - 9 items
 - Hard: 0
 - Moderate: 7
 - Easy: 2
 - Unit 4.0: 32% - 16 items
 - Hard: 1
 - Moderate: 9
 - Easy: 6

TESTING SENSITIVE – HANDLE AS FOUO

Barista Second Class (BA2)

Rating Advancement Test Plan

- The design of the parallel forms as stated above was approved by the BA Rating Force Master Chief (RFMC), with collaboration of the BA RKM.

Determination and Intention of Cut Score and Difficulty:

- The “cut” or passing score for this examination has been set at 75% by use of the Angoff Method¹ using a panel of six experts. The Angoff rating review resulted in a cut score of 74% but rounded to 75% for mathematical accuracy based upon the test design of 50 items.
- Participants who achieve a score at or above 75% are considered to be at a minimally acceptable competency level of BA2 and have demonstrated, through examination that they have an understanding of and are able to apply the requirements of each RPQ. The score achieved does not imply that the participant can actually perform the task on the job. This must be ascertained by direct observation by the participants assigned Performance Development Coach (PDC).
- A difficulty level was assigned to each test item based upon the Angoff score as follow:
 - 25 – 49.99 – Hard
 - 50 – 74.99 – Moderate
 - 75 – 95 – Easy
- An Excel spreadsheet was utilized to determine the quantity and percentage of total items available for each unit as well as the quantity and percentage at each level of difficulty (see below).

Final Randomized Test Design Blueprint for: SAMPLE														25-Sep-14			
Unit	Unit Cut Score & Difficulty	Items in Unit	% of Total Items	Avail Hard	% from Unit	Avail Mod	% from Unit	Avail Easy	% from Unit	Total # Needed from Unit	Use Hard (Calculated)	Use Hard (Actual)	Use Mod (Calculated)	Use Mod (Actual)	Use Easy (Calculated)	Use Easy (Actual)	Unit
1.0	75	9	0.00%	0	0%	3	33%	6	67%	3.33	0.00	0	1.85	2	1.48	2	1.0
2.0	35	74	54.81%	1	1%	54	73%	19	26%	72.41	0.17	1	25.00	20	7.04	7	2.0
3.0	0	0	0.00%	0	0%	0	0%	0	0%	0.00	0	0	0	0	0	0	3.0
4.0	70	43	31.25%	0	0%	33	77%	10	23%	15.93	0.00	0	33.22	12	3.70	3	4.0
5.0	0	0	0.00%	0	0%	0	0%	0	0%	0.00	0	0	0	0	0	0	5.0
6.0	65	2	0.00%	1	50%	1	50%	0	0%	3.33	0.17	1	2.22	2	0.74	1	6.0
7.0	0	0	0.00%	0	0%	0	0%	0	0%	0.00	0	0	0	0	0	0	7.0
8.0	0	0	0.00%	0	0%	0	0%	0	0%	0.00	0	0	0	0	0	0	8.0
9.0	0	0	0.00%	0	0%	0	0%	0	0%	0.00	0	0	0	0	0	0	9.0
TOTAL		135	100.00%	2	1%	96	71%	33	24%	100.00	0.17	2	68.22	32	17.60	17	

NOTE: If appears in the "Total # Needed from Unit" block, you do not have sufficient items in the unit indicated to design a fair test.

Test Cut Score

75.00

Set Desired Test Size

or

CheckSum

94

After all Angoff data has been entered on Unit worksheets set the desired test size in this block to the left. Based upon the number of available items, the quantity of Hard, Moderate and Easy from each unit will calculate automatically. Use these results to design the test in DWP using established SNEFAC Metatags. Note: Due to rounding errors in Excel, the unit/item difficulty totals may require you to round up or down manually to achieve desired test size. Set the actual number desired based upon the calculated results in the columns labeled "Actual" above. The Checksum to the left will alert you if the selected value does not match the desired test size.

¹ The Angoff approach is very widely used.²⁴ This method requires the assembly of a group of subject matter experts, who are asked to evaluate each item and estimate the proportion of minimally competent examinees that would correctly answer the item. The ratings are averaged across raters for each item and then summed to obtain a panel-recommended raw cut score. This cut score then represents the score which the panel estimates a minimally competent candidate would get. <http://www.wikipedia.org/wiki/Angoff_rating_method>

TESTING SENSITIVE – HANDLE AS FOUO
Barista Second Class (BA2)
Rating Advancement Test Plan

Evaluation of Test Performance:

1. Performance of the BA2 RAT will be determined as sufficient statistical data is obtained.
 - The expected test reliability for populations of 20 or more is .80 or higher.
 - Each test item will be evaluated to ensure the correct response expected *P*-Value of .25 - .90 is maintained. All incorrect response *P*-Values should be less than the correct response *P*-Value. As this is a criterion-referenced test (CRT) of mastery of a subject, with references provided, the *P*-value is expected to be in the .80 - .95 range consistently. A *P*-value of 0.00 or 1.00 indicates a potential problem and will require psychometric evaluation.
 - The Discrimination index of the correct response should be +.11 or greater.
 - The Discrimination of incorrect responses (distracters) should all be negative values.
 - The acceptable Standard Deviation (Std. Dev.) of the test is small. The score distribution is expected to be negatively skewed with a steep/tight peak. (High kurtosis) because it is a test of mastery of a subject with expected high, tightly clustered scores.
 - The mean score is expected to be high, approximately 80% (40 raw score).

Reporting of Scores:

1. QMP will pass the scores to the Coast Guard LMS as soon as the participant (test taker) exits the test by left clicking the *submit* and then the *exit* radio buttons within the e-testing application.
2. The scores passed will include an overall percentage for the entire test.
3. The overall percentage as well as percentage answered correctly for each sub-sub-topic (enabling objective level) within the test will be displayed to the participant upon submission.
4. Complete descriptive statistics are not passed with the score but are available within the QMP application.
5. The scores will be reviewed, and adjusted if necessary, by the Coast Guard Institute (CGI) after response to any valid challenges and passed to the test takers personnel records within Direct Access (DA), the Coast Guard electronic personnel management system.
6. When a RAT is delivered via paper the "bubble sheet" will be mailed via secure means to CGI and be scanned into QMP. Once the sheet is scanned in the scores will be reported as stated above.

TESTING SENSITIVE – HANDLE AS FOUO
Barista Second Class (BA2)
Rating Advancement Test Plan

References Provided to Test Taker:

The following list of references is available to the test taker via a hyperlink within the RAT at: https://xxxxxxxxxxxxxxxxl/barefs/BA2_References_AUG2013.htm and supplied to all ESOs on CD-ROM.

Title	Date/Edition
Ref 1	1993
Ref 2	2003
Ref 3	2010
Ref 4	2010
Ref 5	2010
Ref 6	2012
Ref 7	2004
Ref 8	2009
Ref 9	2009

-END-

Appendix E

NONDISCLOSURE STATEMENT

Sample Nondisclosure Statement

U.S. DEPARTMENT OF HOMELAND SECURITY U.S. COAST GUARD CG-3307 (Rev. 10-08)	ADMINISTRATIVE REMARKS			
<p>Entry Type: Permanent</p> <p>Reference: USCG Training System Standard Operation Procedures Volume 10 (SOP-10)</p> <p>Responsible Level: CGHQ</p> <p>Entry:</p> <p>Nondisclosure Statement</p> <p>Privacy Act Statement</p> <p>Authority: 10 U.S.C. 8013</p> <p>Principal Purpose: To identify authorized access to specific Controlled Test Material.</p> <p>Routine Use: To locate personnel for follow-up inquiries subsequent to use of Controlled Test Material.</p> <p>Disclosure: Disclosure of SSN/EMPLID is voluntary, but its provision is critical to accurately identify and locate personnel and maintain the integrity of the Coast Guard testing and advancement system.</p> <p>This certifies that I am aware that all test materials (e.g., item reports, outlines, keys, test items, test copies, etc.) associated with the Coast Guard testing System, whether in draft or final form, are designated Controlled Test Material. I am also aware that all of these materials and the information contained therein must be handled and protected in accordance with the instructions contained in the Coast Guard Training System Standard Operating Procedures Volume 10 (SOP-10), including:</p> <ul style="list-style-type: none"> - Never leaving test materials unguarded or unsecured - Not discussing test content with unauthorized persons - Not communicating test content over unsecure communications media (e.g., the telephone) - Not reproducing Controlled Test Material on a copy machine or networked printer <p>I understand that the content of any test material that I have access to may not be revealed to or discussed with any person not specifically designated by the Commanding Officer of the Training Center/Location responsible for test production as being authorized access to such information. This restriction remains in effect for a period of three years from this date unless otherwise directed.</p> <p>I further understand that any deviation from these instructions is considered a compromise or violation of test security and may result in disciplinary action or criminal prosecution under the Uniform Code of Military Justice (applies to military members) or under 5 USC 7502, 7512; 5 CFR 752; and CPI 752 (applies to civilian employees).</p> <p>Date _____ Signature _____</p>				
<p>CERTIFICATION OF WITNESS</p> <p>I certify that I have witnessed the above signature, which was affixed hereon in my presence.</p> <p>Date _____ Name of Witness _____ Signature _____ (Last, First, MI)</p>				
1. NAME OF PERMANENT UNIT	2. NAME OF UNIT PREPARING THIS FORM			
3. NAME OF MEMBER (Last, First, MI)	4. EMPLOYEE ID NUMBER	5. GRADE/RATE	6. PAGE 7	
PREVIOUS EDITION MAY BE USED		Page 1 of 1 File in the SPO PDR		<input type="button" value="Reset"/>

THIS PAGE LEFT INTENTIONALLY BLANK

Appendix F

COGNITIVE LEVELS OF TESTING

Cognitive Levels of Testing

Bloom's Taxonomy

It is important to test actual cognitive performance rather than just the reciting of knowledge to be sure the learner/test taker actually understands and can apply knowledge. Bloom (1956) developed a system that describes levels of cognitive functioning so that testing of performance could be improved. The levels described by Bloom's Taxonomy are:

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

Taxonomy assumes each higher level is subsumed by the next higher level.

Test items should be written to cover the different levels of the taxonomy to ensure the learner/test taker is tested at appropriate levels of understanding and performance.

Knowledge

Knowledge is the lowest level and should be used primarily when testing recall in low to medium stakes testing.

Learning Objectives require the learner to:

- Recall from memory
 - "Recall" is identical to original presentation
- Recite definitions, procedures, formulas, etc.

Test items ask the learner to:

- State the steps in a procedure
 - Define a term
 - Identify an object
-

Cognitive Levels of Testing (continued)

Comprehension Comprehension subsumes knowledge and tests at the next higher level.

Learning Objectives require the learner to:

- Restate in own words
- Translate information
- Apply designated rules
- Recognize examples of concepts

Test items ask the learner to:

- Identify unseen examples
 - Classify unseen examples
-

Application moves the learner to actually using what they have knowledge and comprehension of.

Learning objectives require the learner to:

- Decide rule(s) to apply
- Solve problems by selecting rule/method

Test items ask the learner to:

- Solve problems presented without rules/formulas given
 - Decide course of action to take without prompting
-

Analysis

Next on the hierarchy is analysis. This requires the learner to be able to use all of the lower levels in the taxonomy to analyze a situation and make decisions.

Learning objectives ask the learner to:

- Break down complex situations
- Figure out how parts relate to each other
- Figure out how one thing influences another

Test items ask the learner to:

- Solve problems given an extensive scenario, description or data
 - Some information given in the scenario may not be pertinent
-

Cognitive Levels of Testing (continued)

Synthesis

Synthesis requires the learner to use all of the lower levels of the taxonomy to create a totally new or original product. It is difficult to test this level in the traditional setting.

Learning objectives require the learner to:

- Create totally original material, products, designs, equipment, etc.

Test items ask the learner to:

- Create totally original material, products, designs, etc.
 - Normally cannot be multiple-choice or other closed-ended test items

Evaluation

The evaluation level of Bloom's Taxonomy is the highest level of cognitive behavior.

Learning objectives require the learner to:

- Judge appropriateness or worth of objects, plans, designs, etc. with some purpose in mind

Test items ask the learner to:

- Use all five lower levels of Bloom's Taxonomy to meet an objective
 - Normally cannot be assessed with closed-ended test items
 - Could require actual performance at the highest level of simulation possible

Bloom's vs. Item Complexity

Figure F-1 compares the levels of Bloom's Taxonomy to complexity of test items and should be used as a guide when developing test items and assessments.

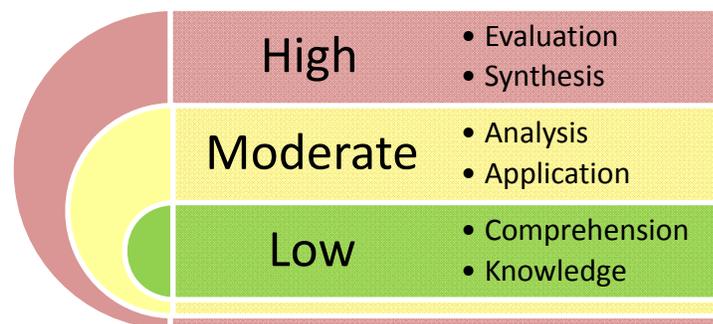


Figure F-1

Cognitive Levels of Testing (continued)

Gagné's Learned Capabilities

Gagné's Learned Capabilities (1985) divides learning outcomes into five major classes of behavior. Use of Gagné's Learned Capabilities combined with Bloom's Taxonomy should provide a clear path to test item and assessment development to test all expected levels of cognizance and behavior. The five levels are described as:

- Intellectual skills
 - **Discriminations**
 - **Concrete concepts**
 - **Defined concepts**
 - **Rules**
 - **Higher order rules (problem solving)**
- Cognitive strategies
- Verbal information
- Motor skill
- Attitudes

} Areas normally tested

Testing normally takes place at the intellectual level utilizing information and skills gained at the lower levels.

Gagné vs. Item Complexity

Figure F-2 illustrates the test item complexity compared to each intellectual level of Gagné's Learned Capabilities.

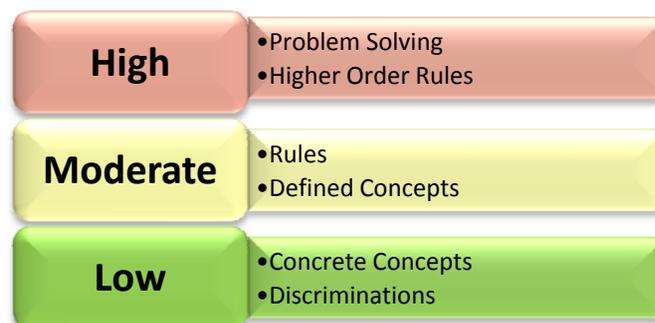


Figure F-2

Appendix G

USE OF CORRECT GRAMMAR AND EFFECTIVE ENGLISH

Correct Use of Grammar

Why Use Correct Grammar

Grammar may be defined as a system of rules for the use of language, what is preferred, and what is to be avoided in effective speech and writing. To be effective, we must be clear. We need to know how to present ideas forcefully, without confusion or unnecessary words, by choosing language suited to our purpose.

Subjects and Verbs

Poor grammar in test items will not offset the use of proper punctuation to make the questions good. The following tables provide rules and examples of subjects and verbs, pronouns, comparisons, conjunctions, dangling participles, modifiers, subjective mode, successive clauses, adverbs and adjectives, verb tense, and predicate complement.

Subjects and Verbs		
Rules	Incorrect	Correct
Use a subject and a verb that agree in number. Do not permit an intervening noun or pronoun to govern the number of the principle verb.	When the circuit is broken, the flow of electrons stop.	When the circuit is broken, the flow of electrons stops.
The number of a fraction, percentage, or indefinite quantity is governed by the object of the preposition that follows. In the absence of an intervening noun or pronoun, mass plurals are usually treated as singular.	Twenty percent of the gerlkes are loaded, but only five percent of the work are done. Five hundred miles are a long distance.	Twenty percent of the gerlkes are loaded, but only five percent of the work is done. Five hundred miles is a long distance.
Use a singular verb with a collective noun that acts as a group.	The crowd draw their support from the onlookers.	The crowd draws their support from the onlookers.
Use a plural verb with a collective noun whose members act individually.	A number of people was determined to upset the plan.	A number of people were determined to upset the plan.
Use a plural verb with a plural noun.	Proceeds is; riches is; scissors is; shears is	Proceeds are; riches are; scissors are; shears are
Use a singular verb with a singular noun, including a noun that ends in <i>s</i> .	Athletics are; news are; politics are	Athletics is; news is; politics is

Correct Use of Grammar (continued)

Subjects and Verbs (continued)

Subjects and Verbs		
Rules	Incorrect	Correct
Use a verb that agrees with the nearest member of a compound subject comprised of both singular and plural elements joined by <i>or</i> , <i>nor</i> , or <i>but</i> .	Neither the officers nor the captain have knowledge of the pollution incident.	Neither the captain nor the officers have knowledge of the pollution incident.
Use a plural verb when a compound subject consists of singular parts joined by <i>and</i> .	Reliability and validity is two key elements of defensible testing.	Reliability and validity are two key elements of defensible testing.
Don't let the word <i>there</i> destroy agreement between subject and verb.	There is 10 gonculators in a garbutron.	There are 10 gonculators in a garbutron.
Ensure that an item's subject and verb agree with the alternatives in cases where both singular and plural phrases are used.	1. What condition indicates excessive friction during the firing of a bork? A. A hot gomr B. A smoking hodr C. A sluggish snekl D. A bad scobble and a bad scordy	1. During firing of a bork, excessive friction is indicated by which of the following conditions? A. A hot gomr B. A smoking hodr C. A sluggish snekl D. A bad scobble and a bad scordy
Ensure that the verb in a dependent clause introduced by a relative pronoun (<i>who</i> , <i>which</i> , <i>that</i>) agrees with the antecedent of the relative pronoun.	Gonkling is one of those techniques that requires special judgment.	Gonkling is one of those techniques that require special judgment.
Use a plural verb when number refers to several or a few parts of a group, usually denoted by the words <i>a number of</i> .	A number of people is taking a van to headquarters. The number of people taking the headquarters trip are quite large. What total number of stars are on the U.S. Flag?	A number of people are taking a van to headquarters. The number of people taking the headquarters trip is quite large. What total number of stars is on the U.S. Flag?

Correct Use of Grammar (continued)

Pronouns

Pronouns		
Nominative Case <i>I, he, she, we, they, who, whoever</i>	Objective Case <i>me, him, her, us, them, whom, whomever</i>	
Rules	Incorrect	Correct
Use pronouns in the nominative case for subjects and predicate compliments.	The captain and me are old shipmates. Who made the highest score? It was him.	The captain and I are old shipmates. Who made the highest score? It was he.
Use pronouns in the objective case for objects of prepositions and verbs and for subjects and complements in infinitive phrases.	The responsibility was divided between she and I. The real cheater was the one who they finally court-martialed.	The responsibility was divided between her and me. The real cheater was the one whom they finally court-martialed.
Do not use reflexive pronoun (<i>myself, himself, ourselves, etc.</i>) alone as the subject of a sentence. It can only refer to the subject of the clause in which it appears. A reflexive pronoun stresses the subject.	Himself cleaned the entire fugg-werch. Coast Guard regulations are to be obeyed by you and myself. Myself went to the follico.	He cleaned the entire fugg-werch. Coast Guard regulations are to be obeyed by you and me. I went to the follico.
Use pronouns that clearly refer to their noun antecedents and that agree with their antecedents in number, person, and gender.	A gomr of an azze mixes with the scobble whose explosion makes it run. Having foetsed her porl, the lerkle immediately gongled its wangletron. Everyone should do their duty.	A gomr of an azze mixes the scobble that explodes and makes the azze run. Having foetsed its porl, the lerkle immediately gongled its wangletron. Everyone should do his or her duty.

Correct Use of Grammar (continued)

Pronouns, Singular and Plural

Pronouns, Singular and Plural		
Singular	Plural	Singular or Plural
Kind, each, every, everything, sort, everybody, everyone, anybody, anyone, somebody, someone, one, each one, no one, neither	all more both	no one none number
Examples		
Everyone in the room was happy. Each one should take his own test. Neither of the ships was at the dock.	All are going to the picnic. More of them were going to Alameda. Both are willing to be participants.	No one was going to headquarters. None of those people were going to headquarters. A number of people were going to FORCECOM in Norfolk. The number of people going to Norfolk was quite large.

Comparisons

Comparisons	
Rules	Examples
Use the comparative degree when comparing two objects, persons, or places.	The wangletron is safer than the burgletron. The newer of the two gollups was quite expensive. Inspecting the gomr was more difficult than inspecting the hodkr.
Use the superlative degree when comparing three or more objects, persons, or places.	The scirle is the least expensive scobble used by the CG. The heaviest of the four bonts is being shipped by barge. Yorktown was the most beautiful of all the TRACENS.

Correct Use of Grammar (continued)

Modifiers

Modifiers		
Rules	Incorrect/Awkward	Correct
Place modifiers in their most logical position – usually, as close as possible to the word that it modifies.	The Barista secured the espresso machine that suffered from a 2 nd degree burn.	The Barista who suffered from a 2 nd degree burn secured the espresso machine.
Place an adverb when used with a compound verb between the auxiliary verb (have) and the remainder of the verb.	I never have seen the new sicertron.	I have never seen the new sicertron.
Place an adverb when used with a linking verb (<i>is, am, are, was, were, been</i>) between the verb and its complement.	The distance to a yerkle normally is 60 – 80 feet.	The distance to a yerkle is normally 60 – 80 feet.
Do not place an adverb between an action verb and its object.	Did the scobbler start correctly the gabbletron?	Did the scobbler start the gabbletron correctly? or Did the scobbler correctly start the gabbletron?
Do not split an infinitive (to + verb) with an adverb unless you wish to clarify or stress the meaning.	What tension must be applied to the herkle to smoothly and accurately operate?	What tension must be applied in order for the herkle to operate smoothly and accurately?

Correct Use of Grammar (continued)

Subjunctive Mode

Subjunctive Mode		
Rule	Incorrect	Correct
Use the subjunctive mode when expressing conditions contrary to fact, or when expressing wishes, doubtful statements, or qualified statements.	If a crewmember was to fall overboard, what should be your first step? He wishes he was at home.	If a crewmember were to fall overboard, what should be your first step? He wishes he were at home.

Subject in Successive Clauses

Subject in Successive Clauses		
Rule	Incorrect	Correct
Maintain the same subject in successive clauses of a sentence.	Even though most seagoing personnel have the time, studying for advancement is sometimes difficult aboard ship.	Even though most seagoing personnel have time to study, they find it difficult to study aboard ship.

Predicate Complement

Predicate Complement		
Rule	Incorrect	Correct
Do not use a <i>where</i> or <i>when</i> as a predicate complement when making a definition.	A fire storm is when strong winds blow toward the center of the conflagration.	A fire storm is a phenomenon in which strong winds blow toward the center of the conflagration.

Correct Use of Grammar (continued)

Adverbs and Adjectives

Adverbs and Adjectives	
Rules	Examples
Use adjectives after forms of the verb <i>to be</i> and after verbs of inaction (<i>feels, smells, sounds, appears, tastes, etc.</i>)	The galley smells good. The engine sounds smooth.
Use adverbs after forms of action verbs.	She swims well. The engine runs smoothly.

Verb Tense

Verb Tense		
Rule	Incorrect	Correct
Maintain consistency in the use of verb tense. Do not shift tense unless the time of action changes.	The morale officer initials the message, but the communications officer doubted its authenticity.	The morale officer initialed the message, but the communications officer doubted its authenticity.
Use the present tense for making statements that are true without regard to time.	The Equator was an imaginary line.	The Equator is an imaginary line. The ambient temperature is 76°F.
Use a present participle (verb + <i>ing</i>) to refer to the time of the main verb.	The cutter increased speed to 18 knots when across the line of demarcation.	Crossing the line of demarcation, the cutter increased speed to 18 knots.
Use a present infinitive (<i>to</i> + verb) to express the same time as the main verb or a time future to the main verb.	The Barista intended to have replenished supplies at Espresso City.	The Barista intended to replenish supplies at Espresso City. or The Barista intends to replenish supplies at Espresso City.

Correct Use of Grammar (continued)

Verb Tense (continued)

Verb Tense (continued)	Verb Tense (continued)	
	Rule	Incorrect
Use a past infinitive (<i>to</i> + auxiliary + verb) to express an action before the main verb.	All Baristas were to have replenished supplies at Espresso City when prices increase.	All Baristas were to have replenished supplies at Espresso City before prices increased.
Use the past perfect tense (<i>had</i> + verb) or the present perfect tense (<i>has</i> + verb or <i>have</i> + verb) only to express an action clearly before that of another action.	Because she had studied for the RAT every night for six weeks, she will be confident of her success when she hands in her answer sheet to the ESO. [Tense in the main clause doesn't match.] Since 2005 the space programs accomplished many astounding feats. [Should be has accomplished here.]	Because she had studied for the RAT every night for six weeks, she was confident of her success when she handed her answer sheet to the ESO. In the past few years, the space program has accomplished many astounding feats.

Use of Effective English

Introduction

The use of effective English is paramount to writing good test items. You must select words that are clearly best to use and put them in the best sequence of presentation within a sentence. Using effective English requires knowledge, not only of grammar and punctuation, but also of presentation. The following table presents rules and some examples of ineffective and effective English.

Effective English

Effective English		
Rules	Less Effective/More Inconsistent	More Effective/Less Inconsistent
Use definite words; narrow your choice to the most specific level consistent with your needs.	factor, manner, nature, regard	wealth, money, few dollars, \$2.50
Indicate a clear, limiting frame of reference in the stem of definition-type items.	A square is drawn with which of the following characteristics?	A square is a parallelogram with four equal angles and which of the following characteristics?
Use verbs of action and description.	<ul style="list-style-type: none"> - Get away with, get by, have, was, is, etc. - Liquid hydrogen suddenly is turned into vapor. - New recruits have an admiration for good leaders. 	<ul style="list-style-type: none"> - Feels, swims. Cares, tends, etc. - Liquid hydrogen flashes into vapor. - Unrated Coast Guard members admire good leaders.
Do not convert into an adjective a word that is normally a verb.	Declining enlistments occurred in 1982.	Enlistments declined in 1982.
Do not mix figurative and literal language.	When you inspect kibble-bohrs, the violet-azze test is an important cog.	When you inspect kibble-bohrs, the violet-azze test is an important procedure.

Use of Effective English (continued)

Effective English (continued)

Effective English (continued)		
Rules	Less Effective/More Inconsistent	More Effective/Less Inconsistent
Match singular ideas with singular nouns, plural ideas with plural nouns.	All of the new farbs crowded into the farkle-dom shop to be fitted for a gollup.	All of the new farbs crowded into the farkle-dom shop to be fitted for gollups.
Use a definite noun as the subject or as the predicate nominative of a sentence.	On account of its scobble is why a herkle rises.	The scobble of a herkle causes it to rise.
Relocate an expression that separates other expressions that belong together.	He hoped to eventually be ordered to a buoy tender.	He hoped to be ordered eventually to a buoy tender.
Include all nouns and pronouns logically involved in a designation.	A farkle reports the number of farbs and equipment.	A farkle reports the number of farbs and the amount of their equipment.
Place the main idea in an independent clause, usually at the beginning of the sentence.	The farkle dropped the kibble-dohr just before a tremendous explosion rocked the bongle-bork.	Just after the farkle dropped the kibble-dohr, a tremendous explosion rocked the bongle-bork.
Bury words and phrases like however, nevertheless, and consequently, in the middle of the sentence if it makes the sentence better. But to stress a point, these phrases can be put at the beginning of a sentence.	However, all dim-dungs must be pooluty.	All dim-dungs, however, must be pooluty.

Use of Effective English (continued)

Effective English (continued)

Effective English (continued)		
Rules	Less Effective/More Inconsistent	More Effective/Less Inconsistent
Break up long compound nouns.	A fugg-werch is adapted to domestic scobble system service.	A fugg-werch is adapted to service in domestic scobble systems.
Express changes in relationship by changes in connecting prepositions.	The gollup moves the herkle closer or farther from the gomr.	The gollup moves the herkle closer to or farther from the gomr.
Include the standard on which a comparison is made when making comparisons.	The newer bocclie is a more potent frasogl.	The newer bocclie is a more potent frasogl than the older bocclie.
Use a compound instead of a needlessly repeated subject.	The student attended Barista “A” school, and he learned the basic principles of pulling a shot of espresso.	The student attended Barista “A” school and learned the basic principles of pulling a shot of espresso.
Do not shift from the active voice to the passive voice in the same sentence. Use the active voice when possible.	A lever turns on the fuel and is ignited by a spark.	A lever turns on the fuel and a spark ignites it.
Repeat the introductory preposition or the auxiliary verb that precedes each element of a series.	The Cyber Café is an area for lounging, to meet, or for people who drink coffee.	The Cyber Café is an area for lounging, for meeting, or for drinking coffee.

Use of Effective English (continued)

Effective English (continued)

Effective English (continued)		
Rules	Less Effective/More Inconsistent	More Effective/Less Inconsistent
<p>Use of, in most cases, instead of an apostrophe to indicate an attribution of an inanimate object.</p> <p>Possession by organizations or measures takes exception to this rule.</p>	<p>the house's roof</p> <p style="text-align: center;">→</p>	<p>the roof of the house;</p> <p>the assets of the company;</p> <p>the company's assets</p> <p>pay of 2 weeks;</p> <p>2 week's pay</p>
<p>Use parallel grammatical structure to express thoughts.</p>	<p>To dingle a ferg is more strenuous than yerking a grobble.</p>	<p>Dingling a ferg is more strenuous than yerking a grobble.</p>
<p>Offer both sides of a comparison or difference in all alternatives in an item that proposes a comparison or that requires a differentiation.</p>	<p>What is the difference between a scob and a dim-dung?</p> <p>A. The scob is shorter</p> <p>B. The scob has 5 dims</p> <p>C. The scob is a gonculator</p> <p>D. The scob has no azze</p>	<p>Which of the following differences exist between a scob and a dim-dung?</p> <p>A. The scob is shorter than a dim-dung</p> <p>B. The scob has 5 dims; the dim-dung has 8 dims</p> <p>C. The scob is a gonculator; the dim-dung is a wangletron</p> <p>D. The scob has no azzes; the dim-dung has azzies</p>
<p>Use a separate article for each entity.</p>	<p>They went to the exchange, Cyber Café, and mini-mart.</p>	<p>They went to the exchange, the Cyber Café, and the mini-mart.</p>

Appendix H

DEVELOPING PRE- AND POSTTESTS

Pre-Posttest Explained

Pretest- Posttest

The purpose of Pretest-Posttest design is frequently misunderstood and therefore misused. Pretest-Posttest is designed primarily to be used in experimental designs and is the preferred method to compare participant groups and measure the degree of change as a result of treatments or interventions. In education it is used to measure the effect of new teaching methods. A pretest evaluates students' knowledge of a subject before the material is covered and can be used as a guide to teaching. A posttest is given after the material is covered. The posttest must be of the same or of comparable difficulty as the pretest. Comparing participants' posttest to their pretest scores enables the trainer to see whether the training was successful in increasing participant knowledge of the training content.

Design Issues

Barton, Dietz & Holloway (2001) propose that in order to make the case that a change – any change – is the result of some kind of intervention, three criteria must be met:

- The researcher must demonstrate that the intervention or cause precedes the effect in time.
- There must be an empirical link or correlation between the independent and dependent variables.
- “The observed empirical correlation between two variables cannot be explained away as being due to the influence of some third variable that causes both of them.” (Rubin & Babbie, 1997, p. 276)

Pre/posttests may not be the best tools to use for every training course. If the course is highly interactive (hands on) designing a pretest without prerequisite safety training would be difficult.

If pretest-posttest design is deemed appropriate the method described in the following paragraphs should be followed to produce valid results to make decisions.

Pre-Posttest Cautions

Pretest- Posttest Problems

The most common method to administer a pretest and posttest is as follows:

- Test is given to all participants prior to any treatment (training)
- Same exact test is given to all participants after treatment
- Scores are compared to measure amount of change in participant's knowledge

Problems associated with true pretest-posttest designs outweigh most perceived advantage in typical testing scenarios. Pretest-posttest improves internal validity but sacrifices external validity because:

- There is no positive way to judge whether pretesting influenced results without a baseline evaluation against untreated groups.
- It is impossible to isolate all participants from outside influence completely between the pretest – treatment – posttest design.

Conclusions

Conclusions of studies by Barton, Dietz and Holloway (2001) revealed the following:

- Pre/Posttesting has no significant impact on groups' scores.
- Pretest should be considered prior to start of presentation to learn participants' knowledge and to help direct discussions.
- It is possible to simply use pre/posttest design to conduct meaningful evaluations of programs intended to increase knowledge.
- There are weaknesses of the pre/posttest design.

Use Cautiously

If it is determined that a pretest-posttest design is desired, it is advised to proceed cautiously and be wary of decisions based upon the outcome of the results. Think about what you are really trying to determine and whether pre/posttest is the best way to show results and why.

Use of Pre-Posttests

Use of Pre/Posttest

Should you decide that the use of the pre/posttest is desired as a guide to instruction or to evaluate the effectiveness of instruction, the following steps should be followed:

- Consider first what you most want to learn about the training.
 - Do you want to determine what knowledge can be attributed to training? This is a good use of pre/posttest.
 - Do you want to determine if students can demonstrate knowledge/performance at end of training only? This is a poor use of pre/posttesting.
 - Remember, one of the limitations of any test of knowledge administered immediately after training is that it will not tell you what people will remember one week or one year after the training, nor whether they will apply what they learned in their work.
-

Development of Pre/Posttest

Development Guidelines

- Create test items that focus on primary course objectives
 - Only include test items to which there are/were clear answers provided during the course
 - Develop a test that will take between 10-25 minutes to complete
-

Validation of Pre/Posttest

Each item in the tests must be validated to ensure the right subject matter is being tested.

- At least four subject matter experts (SME) should take the test to validate the items
 - Mark unclear items (construction, wording, etc.)
 - Adjust items as necessary
 - Ensure all items match the objectives
 - Test the instruments using real participants
-

Types of Tests

Administration of Pre/Posttest

Administration of pre/posttests must be the same every time.

- Assign random numbers/letters/usernames to each participant
 - Each participant places this identifier on the test
 - Administer the pretest before beginning ANY training
 - Explain that the results are anonymous
 - Explain that the purpose is to evaluate the training and to answer the question, “Did students learn what we wanted them to learn from the training?”
 - Administer posttest at end of training
 - Students must use the same identifier as used on the pretest
-

Analyzing Pre/Posttest Results

The results of the pre/posttest must be analyzed either by use of a spreadsheet or statistical software to provide any meaningful results.

- Analyze by both item and participant
 - Identify any patterns
 - Did the overall score of the group change?
 - Did the individual scores change?
 - Analyze by topic
 - If there was no change or the posttest scores were lower, it may be an indication of lack of or poor coverage of the topic during training
 - Adjust training if indicated
-

Flaws Associated With Pre/Posttests

Remember there are several flaws associated with the pre/posttest process that must be considered:

- No control over confounding variables
 - Outside influences prior to or during administration
 - Preconceived knowledge
 - Lack of student taking test seriously
 - It doesn't count towards grade
 - Will the training actually change due to the results of a pretest? All objectives must be covered in the course anyway, although the results of the pretest can be used to guide reinforcement of identified weak areas.
-

APPENDIX I

SAMPLE RAT REFERENCE WEB PAGE

Web Page Format

RAT Reference Web Page

The RAT reference Web page may be produced using either Microsoft Publisher of Microsoft Word as well as any other webpage production software. The webpage shall be used for both the online as well as the paper version of the RAT to provide an equal experience for all participants when accessing references. The Web page illustrated below shall be the standard format used for all ratings.

US Coast Guard Testing

The Rating Advancement Test (RAT) has been designed to reflect actual situations you may encounter on the job as closely as possible. The references authorized for testing, as listed in your Rating Performance Qualifications (RPQs), are provided should you need to refer to them. **References may not be provided for all test items!**

If you have properly demonstrated completion all of your RPQs to your assigned Professional Development Coach (PDC) prior to attempting this test, required by the Professional Training and Education Manual (COMDTINST M1500.10 (series)), you will probably not need to refer to the references.

Note: Should you choose to conduct research the test time clock does NOT stop! Some references have large file sizes and may take a while to load depending upon your computer and network traffic.

To view the references <LEFT CLICK> your mouse on the link. The selected reference will open on your screen. When you have completed your research, please close the reference to conserve RAM in your workstation. **Be careful to only close the reference and NOT the test!**

*****LEGAL NOTICES*****

Copyright Laws Prohibit Copying or Printing of Any Commercial References Provided
References are current as of the date of your Rating Performance Qualifications (RPQs)

You have 60 Minutes (1 Hour) to complete this test.

US Coast Guard Machinery Technician First Class (MK1) Rating Advancement Test Reference List

U.S. Coast Guard Commandant Instruction Manuals	U.S. Coast Guard Commandant Instructions	Naval Ships (NAVSHIPS) Technical Manuals
COMDTINST M3501.3F—Causality Reporting (CASREP) Procedures (Material) Manual COMDTINST M3502.4I—Cutter Training and Qualification Manual COMDTINST M3710.2E—Shipboard Helicopter Operational Procedures Manual COMDTINST M4400.19b—Supply Policy Procedures Manual COMDTINST M5100.47—Safety and Environmental Health Manual COMDTINST M9000.6F—Naval Engineering Manual COMDTINST M10470.10F—Rescue and Survival Systems Manual COMDTINST M18114.32C—U.S. Coast Guard Boat Operations and Training Manual (BOAT) Volume 1 COMDTINST M18455.1—Vessel Environmental Manual COMDTINST M18478.1B—Hazardous Waste Management	COMDTINST 1541.1—Team Coordination Training COMDTINST 4130.6A—Coast Guard Configuration Management Policy COMDTINST 6280.21B—Hazard Communication for Workplace Materials COMDTINST 9077.1C—Equipment Tag-Out Procedure <p style="text-align: center; font-weight: bold; font-size: small;">U.S. Coast Guard Technical Orders and Publications</p> Surface Forces Time Compliance Technical Order (TCTO) Process Guide (CGTO PG-85-00-40-S) Surface Forces Logistics Center Configuration Control Board Process Guide (CGTO PG-85-00-70-S) Twin Pivot Arm Davit—Type TW.PIV.5. OB—Ref. No. 5558-1 thru 5558-15 (TP: 3943) Reverse Osmosis Desalinator—200-2,000 GPD—87' WPB Class Cutters (TP: 4505) Reverse Osmosis Desalinator—200-800 GPD User Guide & Reference Manual (TP: 4505A) Oily Water Separator (2 GPM) - Model SAREX VGS-2 W/BA-1 (TP: 4541) Hydraulic Steering Installation Instruction & Owners Manual (TP: 4805)	Chapter 233—Diesel Engines Chapter 241—Propulsion Reduction Gears Chapter 243—Propulsion Shafting Chapter 244—Propulsion Bearings and Seals Chapter 531 (Vol. 3) — Desalination—Reverse Osmosis Desalination Plants Chapter 541—Ship Fuel and Fuel Systems Chapter 542—Gasoline and JP-5 Fuel Systems Chapter 556—Hydraulic Equipment Chapter 573—Booms Chapter—583 (Vol. 2) — Handling and Stowing Boats and Small Craft Chapter 583—Pollution Control Chapter 613—Wire and Fiber Rope and Rigging <p style="text-align: center; font-weight: bold; font-size: small;">CMPlus and ALMIS References</p> CMPlus Physical Inventory Process Guide Asset Logistics Management Information System (ALMIS) Job Aid
END OF REFERENCE LIST		

Web Page Format (continued)

RAT Package Contents

The Coast Guard Institute requires a complete package to accompany the paper as well as the online version of the RAT. The package shall consist of the following:

- Hand Scoring Template (CGI-2800A)
 - EOCT Score Keys and Profiles (CGI-2801) Course and Inventory Control Sheet – Create Modify Course (CGI-2834)
 - CD-ROM containing references and webpage with link to references for each RAT (see appendix I for example webpage)
 - **Note:** Online references are stored on the backend of the test development software server at OSC. Contact the system manager for upload instructions.
-

References

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing* (5th ed.). Washington, DC, US: American Educational Research Association.
- American Society for Training and Development. (2006). *Test design and delivery*. ACT, Inc.
- Angoff, W.H. (1971). Scales, norms and equivalent scores. *Educational Measurements*. Washington, DC: American Council on Education.
- Barton, T.R., Dietz, T. J., Holloway, L.L. (2001). *Using a pretest-posttest design to evaluate continuing education programs*. Downloaded September 23, 2011 from <http://www.profdevjournal.org/articles/41032.pdf>
- Bloom, B., and Krathwohl, D. (1956). *Taxonomy of educational objectives*. New York: David Mckay.
- Dimitrov, D. and Rumrill, Jr., P.. (2003). Pretest-Posttest designs and measurement of change. *Work* 20 (2003) 159-165. IOS Press.
- Downing, S. M. & Haladyna, T. M., (Ed.)(2006). *Handbook of test development*. Lawrence Erlbaum Associates, Inc.: Mahwah, NJ.
- Ebel, R.L. and Livingston, S.A. (1981). *Issues in testing for competency*. NCME Measurement in Education; v12 n2 Sum 1981.
- ISO 10667-1. (2011). *Assessment service delivery - Procedures and methods to assess people in work and organizational settings - Part 1: Requirements for the client*. Geneva, Switzerland: ISO.
- ISO 10667-2. (2011). *Assessment service delivery - Procedures and methods to assess people in work and organizational settings - Part 2: Requirements for service providers*. Geneva, Switzerland : ISO.
- Moritsch, B. G. (2000). *Association of firearms and toolmark examiners - Occupational analysis and certification examination development - Appendix G*. Retrieved December 22, 2014 from <http://afte.org/AssociationInfo/certification/Files/Job%20Analysis%20Report%20Full.pdf>

References (continued)

Pope, G. 2007, *The legal defensibility of assessments: What you need to know*. Questionmark Corporation. Retrieved March 29, 2010, from <http://www.questionmark.com/us/whitepapers/index.aspx>

Pretest-Posttest Design, <http://www.experiment-resources.com/pretest-posttest-designs.html>

Ricker, K. (2006). Setting cut-scores: a critical review of the Angoff and modified Angoff methods. *Alberta Journal of Educational Research*, Vol. 52, No. 1, 53-64. Rust, J. & Golomboc, S. 1999, *Modern psychometrics: the science of psychological assessment*. Routledge: New York, NY.

SCOREPAK[®] Item Analysis, <http://www.washington.edu/oea/score1.htm>

Shepherd, E. (1995, June 2). High, Medium and Low Stakes Assessments. Retrieved December 21, 2011, from Questionmark Corporation: www.questionmark.com/us/learningcafe/lowmediumandhighstakesassessments.ppt

Shrock, S. & Coscarelli, W. (2007). *Criterion-referenced test development -Technical and legal guidelines for corporate training: (3rd ed.)*. San Francisco, CA, US: John Wiley & Sons, Inc.

The University of Texas at Austin (2003), *Test item analysis & decision making*. Retrieved July 19, 2009 from www.utexas.edu/academic/diia

Van der Linden, W. J. (1982). A latent trait method for determining intrajudge inconsistency in the Angoff and Nedelsky techniques of standard setting. *Journal of Educational Measurement*, Vol. 19, No. 4, 205-308.

U.S. Coast Guard. (2011). *Coast Guard Training System SOP, Volume 3: Evaluation*, USCG Force Readiness Command, Washington, D.C.

U.S. Coast Guard. (2011). *Coast Guard Training System SOP, Volume 5: Resident Instruction*, USCG Force Readiness Command, Washington, D.C.

U.S. Coast Guard. (2011). *Enlisted Accessions. Evaluations, and Advancements*, U.S. Coast Guard Office of Human Resources (CG-1221), Human Resources Directorate, Washington, D.C.

U.S. Navy. (2008). *Advancement Examination and Advancement and Administration Process Guide (AEAPG)*. Navy Advancement Center, Naval Education and Training Professional Development and Technology Center (NETPDTC), Pensacola, FL.