NAVY SCHOOL HOUSE TESTING MANAGEMENT MANUAL


DISTRIBUTION STATEMENT A: Approved for public release; distribution is unlimited.
LETTER OF PROMULGATION FOR NADEVTRA 132

1. This guidance manual has been extensively revised. Most of the revisions are in response to user comments and reflect a continuing effort to increase the manual’s utility to the training field. NADEVTRA 132A supersedes and replaces NAVEDTRA 132.

2. The procedures in this manual management of testing procedures at Navy School under NETC purview and their roles in the Organization Structure, management of staff, student and curriculum. In addition, an understanding of assessment strategies and support functions. This manual is intended for use by military, civil service, and contractor personnel engaged in Navy training materials development and modification.

3. This publication is available electronically at: Navy Knowledge Online (NKO) - NETC N74 Learning Standards Homepage; and Navy Marine Corps Intranet's (NMCI) Total Records and Information Management (TRIM).

4. Corrections and comments concerning this manual are invited and should be addressed to the Naval Education and Training Command, attention: N7.

5. Reviewed and approved.

/s/
J. F. KILKENNY

DRAFT

PUBLISHED BY DIRECTION OF
COMMANDER NAVAL EDUCATION AND TRAINING COMMAND

NAVEDTRA 132A
### RECORD OF CHANGE

<table>
<thead>
<tr>
<th>Number Description of Change</th>
<th>Entered by</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FORWARD

NAVEDTRA SERIES MANUALS:

- NAVEDTRA 130 - Task Based Curriculum Development Manual
- NAVEDTRA 131 - Personnel Performance Profile Based Curriculum Development Manual
- NAVEDTRA 134 - Navy Instructor Manual
- NAVEDTRA 135 - Navy School Management Manual

The NAVEDTRA 130 series of manuals provides fundamental guidance, within the Naval Education and Training Command (NETC), for the development of curricula, the delivery of instruction, and the management and evaluation of training programs.

These manuals do not supersede the directive policy established by Commander, Naval Education, and Training Command Instructions (NETCINSTs) in these subject areas. Rather, they supplement the NETCINSTs in two important ways. First, they reflect the philosophical principles underlying NETC policy for curriculum, instruction, and evaluation and second, they provide procedures for carrying out that policy.

Each of the NAVEDTRA 130 series manuals is designed as a stand-alone document to serve a specific user group such as curriculum developers, instructors, training managers, or evaluators of training. The manuals are, however, interrelated and cross-referenced to one another.

SCOPE

The NAVEDTRA 132 (series), Navy School House Testing Management Manual, outlines the Naval Education and Training Command (NETC) requirement for all training activities to conduct quality testing programs. The testing program is design to measure the attainment of desired learning outcomes. A quality testing program will ensure that successful course graduates have achieved at least the minimum level of skill and knowledge necessary to do the job for which they were trained. It will also identify those students who have failed to achieve
the desired level of skill and knowledge and need further training. The methods used to measure course objectives will impact significantly on the quality of the testing program.

The purpose of this handbook is to provide information and guidance in the development, management and analysis of testing programs throughout NETC. Although no references are cited, testing literature has been reviewed. Concepts and guidelines pertaining to quality testing programs outline in this handbook are not necessarily the ideas and theories of the authors but are shared opinions of many authors of testing publications considered expert in evaluation strategies.

The guidelines set forth in this handbook are not intended to conflict with any higher level authority testing policies or procedures. In those cases where there appears to be conflict or disagreement, please notify the NETC point of contact. We solicit you comments and recommendation for improvement. The NETC point of contact is Dr. Larry Reaves, N745B, DSN 262-9871.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TITLE PAGE</td>
<td>I</td>
</tr>
<tr>
<td>LETTER OF PROMULGATION</td>
<td>II</td>
</tr>
<tr>
<td>CHANGE RECORD</td>
<td>III</td>
</tr>
<tr>
<td>FOREWORD</td>
<td>IV</td>
</tr>
<tr>
<td>TABLE OF CONTENT</td>
<td>VI</td>
</tr>
<tr>
<td>LIST OF ACRONYMS</td>
<td>VIII</td>
</tr>
<tr>
<td>CHAPTER 1 - INTRODUCTION TO TESTING</td>
<td>1-1</td>
</tr>
<tr>
<td>CHAPTER 2 - TEST DESIGN</td>
<td>2-1</td>
</tr>
<tr>
<td>CHAPTER 3 - PERFORMANCE TEST DEVELOPMENT</td>
<td>3-1</td>
</tr>
<tr>
<td>CHAPTER 4 - KNOWLEDGE TEST DESIGN</td>
<td>4-1</td>
</tr>
<tr>
<td>CHAPTER 5 - TEST ITEM CONSTRUCTION</td>
<td>5-1</td>
</tr>
<tr>
<td>CHAPTER 6 - KNOWLEDGE TEST DEVELOPMENT</td>
<td>6-1</td>
</tr>
<tr>
<td>CHAPTER 7 - MANAGING A TEST PROGRAM</td>
<td>7-1</td>
</tr>
<tr>
<td>CHAPTER 8 - TEST AND TEST ITEM ANALYSIS</td>
<td>8-1</td>
</tr>
<tr>
<td>CHAPTER 9 - TESTING PLAN</td>
<td>9-1</td>
</tr>
<tr>
<td>APPENDICES’</td>
<td></td>
</tr>
<tr>
<td>A - TERMS AND DEFINITIONS</td>
<td>A-1</td>
</tr>
<tr>
<td>B - SAMPLE CHECKLIST AND RATING SCALES</td>
<td>B-1</td>
</tr>
<tr>
<td>C - SAMPLE TEST/TEST ITEM VALIDATION WORKSHEETS</td>
<td>C-1</td>
</tr>
<tr>
<td>D - SAMPLE TESTING PLAN</td>
<td>D-1</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIGURES’</td>
<td></td>
</tr>
<tr>
<td>2-1 – ANALYZE OBJECTIVES FOR CRITICALITY</td>
<td>2–7</td>
</tr>
<tr>
<td>4-1 – SAMPLE TEST DESIGN FOR A PARTICULAR TEST</td>
<td>4-1</td>
</tr>
<tr>
<td>5-1 – COMPARISON OF TEST ITEMS TO LEARNING LEVEL CHART</td>
<td>5–21</td>
</tr>
<tr>
<td>6-1 – COMPLETED TEST DESIGN SUMMARY</td>
<td>6-6</td>
</tr>
<tr>
<td>8-1 – ITEM DIFFICULTY FOR ESSAY ITEMS</td>
<td>8-8</td>
</tr>
</tbody>
</table>
### LIST OF ACRONYMS

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARB</td>
<td>Academic Review Boards</td>
</tr>
<tr>
<td>CANTRAC</td>
<td>Catalog of Navy Training Courses</td>
</tr>
<tr>
<td>CCA</td>
<td>Curriculum Control Authority</td>
</tr>
<tr>
<td>CCM</td>
<td>Course Curriculum Model Manager</td>
</tr>
<tr>
<td>CDP</td>
<td>Course Data Processing</td>
</tr>
<tr>
<td>CIN</td>
<td>Course Identification Number</td>
</tr>
<tr>
<td>CMS</td>
<td>Course Master Schedule (per NAVEDTRA 130/131)</td>
</tr>
<tr>
<td>FCR</td>
<td>Formal Course Review</td>
</tr>
<tr>
<td>FCA</td>
<td>Fleet Concentration Area</td>
</tr>
<tr>
<td>HPPR</td>
<td>Human Performance Requirements Review</td>
</tr>
<tr>
<td>IS</td>
<td>In-Service Training</td>
</tr>
<tr>
<td>ISS</td>
<td>Instructional Systems Specialists</td>
</tr>
<tr>
<td>ITRO</td>
<td>Inter-Service Training Review Organization</td>
</tr>
<tr>
<td>JDTA</td>
<td>Job Duty Task Analysis</td>
</tr>
<tr>
<td>JIT</td>
<td>Journeyman Instructor Training</td>
</tr>
<tr>
<td>KO</td>
<td>Knowledge Objectives</td>
</tr>
<tr>
<td>LC</td>
<td>Learning Center</td>
</tr>
<tr>
<td>LS</td>
<td>Learning Site</td>
</tr>
<tr>
<td>LSO</td>
<td>Learning Standards Officer</td>
</tr>
<tr>
<td>LP</td>
<td>Lesson Plan</td>
</tr>
<tr>
<td>MCS</td>
<td>Master Course Schedule (per CeTARS)</td>
</tr>
<tr>
<td>NEC</td>
<td>Navy Enlisted Classification</td>
</tr>
<tr>
<td>NETC</td>
<td>Naval Education and Training Command (Staff)</td>
</tr>
<tr>
<td>NETPDTC</td>
<td>Naval Education Training Professional Development and Technology Center</td>
</tr>
<tr>
<td>NKO</td>
<td>Navy Knowledge On-Line</td>
</tr>
<tr>
<td>NTMPS</td>
<td>Navy Training Management &amp; Planning System</td>
</tr>
<tr>
<td>NSTC</td>
<td>Naval Service Training Command</td>
</tr>
<tr>
<td>NTSP</td>
<td>Navy Training System Plan</td>
</tr>
<tr>
<td>OJT</td>
<td>On-The-Job Training</td>
</tr>
<tr>
<td>OIC</td>
<td>Officer-In-Charge</td>
</tr>
<tr>
<td>SIP</td>
<td>Student Input Plan</td>
</tr>
<tr>
<td>SME</td>
<td>Subject Matter Expert</td>
</tr>
<tr>
<td>SMF</td>
<td>Student Master File</td>
</tr>
<tr>
<td>SOW</td>
<td>Statement of Work</td>
</tr>
<tr>
<td>TCCD</td>
<td>Training Course Control Documents</td>
</tr>
<tr>
<td>TO</td>
<td>Testing Officer</td>
</tr>
<tr>
<td>TPP</td>
<td>Training Project Plan</td>
</tr>
<tr>
<td>TQI</td>
<td>Training Quality Indicator</td>
</tr>
<tr>
<td>TS</td>
<td>Training Specialist</td>
</tr>
<tr>
<td>TSA</td>
<td>Training Support Agency</td>
</tr>
</tbody>
</table>
CHAPTER ONE

INTRODUCTION TO TESTING
INTRODUCTION

The NAVEDTRA 132A outlines NETC'S requirement for a testing program. The testing program, or the way objectives are measured, is one of the best methods available to assess whether the graduates have obtained the knowledge and skills necessary to do the job. This chapter is introductory in nature and will provide a general overview of the purpose of a testing program and the different types of tools that may be used to measure student achievement. Appendix A provides an overview of the terms and definitions used throughout this manual.

- Purpose of a Testing Program
- Ways of Measuring Objectives

SECTION 1 - PURPOSE OF TESTING PROGRAM

1.1. The primary purpose of a testing program is to measure student achievement of knowledge and skill objectives. This may be done in any or all of the following ways: performance tests, knowledge tests, practical work, and inference.

- Performance tests are sample work situations in which students demonstrate the ability to complete a task or job. Since the goal of many courses is to train students to perform a skill, performance testing may constitute a significant portion of the testing conducted in a course. Courses with skill objectives will measure the student's accomplishment of the objectives either through practical work or through performance testing.

- Knowledge tests are used to measure a student's ability to recognize facts, recall information, comprehend principles, evaluate functions or concepts and analyze findings or results. Knowledge tests are important in technical training because they measure a student's ability to understand knowledge in support of the performance of a skill. For example, in the process of controlling an aircraft, air traffic controllers are required to interpret signals as seen on a radarscope. To ensure that this performance is accomplished in an efficient and safe
manner, the student's ability to recognize the signals and recall what each represents should be measured through knowledge testing prior to the performance test.

- **Knowledge tests are either open or closed book.** Open book tests are used to evaluate a student's ability to locate and record items using references. Open book tests are used when the knowledge supporting a skill performed on the job requires the use of documentation. When the student is required to memorize the knowledge or if the knowledge tested is normally required on the job without reference to documentation, a closed book test should be used.

- **Practical work** is work done in the lab, on the line, or in the field, with the understanding that a classroom becomes a lab in cases where training involves "paper and pencil" activities. Practical work may be used to measure skill and KOs. A lab exercise that allows the student to practice a particular skill or group of skills prior to performance testing is an example of practical work. Homework assignments, quizzes, or problem sheets are all examples of practical work that may be used to measure KOs.

- **Inference** is the process of measuring some objectives through the measurement of higher-level objectives. For example, a KO requiring students to label the parts of an oscilloscope may be measured during the performance test that measures a skill objective requiring students to use the oscilloscope when troubleshooting a faulty component.

- **Courses with knowledge safety objectives** will have a program in place to ensure that all knowledge safety objectives are measured satisfactory prior to performance in a laboratory situation. This may be accomplished through oral tests, written tests, quizzes, etc.

1.2. The second purpose of a testing program overlaps the first in that the program should be specifically designed to assess a student's ability to reason and understand theory concepts in support of skill performance. Each time a knowledge test is planned, consider the skill objective. What information must students know in order to perform the skill? These objectives should be measured through knowledge testing.
Tests are designed to identify, early in the course, students who are failing to attain the objectives. These students are then provided additional training in these areas. Therefore, tests should be administered frequently enough, and at important intervals in training, in order to identify these students. The testing program should be viewed as a method to help student's acquire the knowledge and skills necessary to do a job not a method to "weed out" those who can't make it. But if after additional training, students are still unable to grasp the material, they may be attrited. It is important to attrite students only after every effort has been made to help them understand the material.

Testing programs also serve the purpose of providing feedback to the students as to their progress in attaining the objectives. After each test is administered and graded, students should be informed as to their progress. This is usually done by a review of the test followed by remediation if necessary. More information on how to use tests to provide feedback is provided in Chapter 7, Testing Program Management.

If the instructors and course managers approach the testing program with a positive attitude, it can be used to motivate students to learn and to reinforce the knowledge and skills taught. An instructor with a positive attitude will use the tests to identify areas in which students can improve; not use them to make students feel like failures.

The last purpose of the testing program is to evaluate instructor and curriculum effectiveness and to improve the instructional program. This is accomplished through the analysis program that is discussed in Chapter 8, Test and Test Item Analysis.

SECTION 2 - FUNCTIONS OF A TESTING PROGRAM

2.1. Testing programs are divided into three functions: test development, test management and test/test item analysis. The testing program is an integrated program, divided into separate functions that overlap and are continuous. Not only do the functions overlap as to when they are accomplished, the responsibility for accomplishing these functions is shared by
the Learning Standards Office (LSO), course managers, instructors, and curriculum developers.

- The curriculum developer normally accomplishes test development during a development/revision project. During the process, the curriculum developer designs a testing program that measures the objectives, develops performance tests and grading criteria, writes knowledge test items, develops knowledge tests with criteria to determine minimum passing grade and validates all of the above. Test development may also occur when it becomes necessary to add to or revise an existing test item bank.

- Test management involves managing the complete program after the tests are determined to be valid. This process includes establishing procedures for test material security, test administration, test grading, test review, remediation, and retesting. Once the procedures are established, the testing managers are responsible for ensuring each is applied in an effective and efficient manner.

- Test and test item analysis includes conducting the statistical analysis of tests and test items after the tests are administered. Testing officers and the LSO normally performs this function.

SECTION 3 - METHODS OF DETERMINING A GRADE

3.1. As already discussed, it is necessary to decide how the student's achievement of the objectives will be measured. It is also necessary to determine how student achievement will be expressed in the form of a grade. The purpose of the grading process is to communicate to the student, the fleet and the instructor whether the student has successfully completed the objectives and, in some instances, how well the student has achieved the objectives. There are two acceptable methods of determining a grade: numerical grading and satisfactory/unsatisfactory grading.

- Numerical grades are used when a task can be accomplished with varying degrees of performance. For example, a Yeoman typing a report for a performance test may complete it with no errors or it may contain several errors. Assigning a
numerical grade based on the number of errors in the performance test identifies varying degrees of student performance.

- Numerical grades are used when it is necessary to differentiate levels of achievement between students. For example, when class standing is required or is used as a motivational incentive to reward the top students, numerical grades should be used.

- Finally, numerical grades are used to provide feedback to students as to how well they are doing in relation to the class as a whole. Numerical grades allow for specific feedback and ease of explanation of the grade.

3.2. In order to accomplish standardization in the area of numerical grades, a grading scale has been established. This scale is designed to provide a uniform understanding, both to the student and to the fleet; of the grade a student is assigned. The grades do not represent a percentage rather a placement on a scale. This scale applies to both knowledge and performance testing. The following is the interpretation of the scale:

- **90-100** - *Superior* understanding. Graduates are able to perform quickly and efficiently with little or no supervision. For example, when an "A" is earned in school, the student has attained a superior understanding of the material.

- **80-89** - *Above average* understanding. Graduates perform efficiently with little supervision. Above average understanding equates to a "B".

- **70-79** - *Average* understanding. Graduates complete assignments with minor errors. Supervision is required. Average understanding equates to a "C".

- **63-69** - *Minimum* understanding. Graduates cannot complete assignments without additional instruction and are normally completed with increased supervision. Minimum understanding equates to a "D".

- **0-62** - *Inferior* understanding. Students are unable to meet standards. Inferior understanding equates to an "F".

- **Satisfactory/Unsatisfactory Grading (SAT/UNSAT)** is a method of grading when the performance of a skill or application
of knowledge is either accomplished or not accomplished with no varying degrees of performance. For example, an Aviation Ordnanceman either loads the bomb on the aircraft successfully or unsuccessfully; there is no marginal or outstanding performance. Electronics technicians must know all, not just some, of the safety precautions before working on equipment. When SAT/UNSAT grading is used it does not differentiate between the outstanding student and the minimum student. A minimum requirement is established and all students must meet that minimum.

3.3. Minimum Passing Grade for a Course reflects the minimum acceptable understanding/performance required of the graduate to perform the job in the fleet or follow-on training. The minimum passing grade for a course may be based on the numerical method or the SAT/UNSAT method.

- When the numerical method is used, the minimum passing grade is based on the grading scale discussed on page 6.
- When the SAT/UNSAT method is used, minimum acceptable overall performance must be identified. For example, to satisfactorily complete an equipment maintenance course, the technician must successfully identify 8 of 10 faulty components on the series of troubleshooting performance tests.
- The minimum passing grade for a course is determined by the Course Curriculum Model Manager (CCMM) and approved by the Course Curriculum Authority (CCA) and shall convey the minimum passing grade to the requirements sponsor for their approval. Anytime a minimum passing grade is established at a number higher than 63 or when a SAT/UNSAT method is used, a justification will be provided in the course-testing plan.
- When determining the minimum passing grade, several factors should be taken into consideration. One factor is the criticality or importance of the objectives. For example, if the graduate is expected to perform critical tasks immediately upon arrival to the fleet, then a grade higher than minimum might be appropriate.
- If the graduate will be performing tasks with little or no supervision, then a grade of 75 or above might be appropriate.
If the graduate will be attending a follow-on course of instruction, then the minimum passing grade should be based on the prerequisite knowledge and skill of the follow-on training.

If the graduate will receive extensive On the Job Training (OJT) after graduation or will be performing with supervision, then 63 might be appropriate.

The important fact to remember is the minimum passing grade should be based on the level of performance required of the student upon graduation. While it might be desirable to graduate students with an above average understanding, it may not be practical or necessary.

SECTION 4 - TYPES OF TESTS

There are five different types of tests that can be used in training: pretest, quiz, progress test, comprehensive test and oral test. Each has a different purpose and will be discussed in the following paragraphs.

A pretest is a test administered at the beginning of the course and prior to any instruction. It may be used in the following ways:

4.1. Validation of Material. When a test is administered both at the beginning of the course and again at the end it is normally being used to validate a new course of instruction. A comparison of the results of the two tests helps determine the effectiveness of the instruction. For example, the results of the first test may indicate students do not understand the theory of operation of a transistor (50 of the 60 students answered items on transistors incorrectly). After two weeks of instruction, the same test is administered. The results indicate that 50 of the 60 students answered the items on transistors correctly. Based on these results, the assumption is that learning took place. This type of testing may also be used anytime new material is introduced into a course.

4.2. Acceleration. Pre-tests may be used to determine a student's potential for acceleration through a course or unit of instruction. The pretest is similar to the existing course or unit test and is designed to measure mastery of the learning
objectives. For example, if students administered a pre-test answer 24 of 25 algebra items correctly, it may be cost effective to accelerate the students through this portion of training and move them on to the next unit or module.

4.3. Remediation. Pre-tests may be used to determine a student's need for remediation. This type of pretesting measures prerequisite knowledge and skills necessary to meet entry-level requirements. For example, if a student is required to have knowledge of electricity and electronics prior to attending Fire control Technician Class "A" school, a pre-test may be administered to determine if the student possesses the prerequisite knowledge. If the student does not demonstrate the required understanding, a remediation program can be established to prevent difficulties later in training. This type of pre-test will not be used to "weed out" students before, or early in, training.

4.4. A Quiz is a short test used by the instructor to measure achievement of recently taught material. The quiz may be given as often as desired and may or may not be a part of the student's grade. If used in determining a student's grade, quizzes and testing procedures must be standardized. If not, the instructor may prepare/administer the quiz within guidelines established by the course managers. The quiz is normally not retested. If quizzes are used for grading purposes, they will be considered a part of the practical work grade.

4.5. A Progress Test measures knowledge and/or performance. The results are used to determine how a student is progressing toward the achievement of the objectives. The progress test measures lower level objectives or objectives that are required in order to accomplish the terminal or course objectives. A progress test is normally administered for every 40-50 periods of instructional material. How often a test is given may vary based on the complexity of the material. For example, it may be necessary to measure complex material in small blocks of information. In this case, progress tests may be administered more frequently than once a week. On the other hand, administering tests too frequently may cause the student to rely on short-term memorization and not comprehension.
4.6. A Comprehensive Test may be given at the end of instruction or after large blocks of material to measure mastery of the critical objectives or retention of critical information previously measured. Comprehensive tests may measure performance, knowledge or both. When determining what type of test to use, consider the objectives or the desired outcomes. What must students know or be able to do immediately upon graduation? The answer to this question will determine the type(s) of comprehensive tests and the information to be tested. Two types are discussed on the next page.

- **Within-course comprehensive tests** are administered for courses that are longer in length and/or are of a more complex nature where it may not be practical to administer one final test. For example, if a course is 10 weeks long, it may be appropriate to administer two within course comprehensive tests rather than a final. In any case, the within course comprehensive tests must measure mastery of critical objectives or retention of material previously measured.

- **Final comprehensive tests** are given at the end of the course and once again must measure mastery and/or retention of the critical objectives. Courses are required to administer either a final comprehensive test or within-course comprehensive tests.

4.7. An Oral Test is normally administered in a board type situation. This means the students are asked to respond to questions orally in front of a panel of evaluators. Oral tests are best used when performance of the job in the fleet requires a verbal demonstration of a skill. The following guidelines should be considered when determining the need for oral testing.

- **Student Instructor Ratio/Class Size.** If the student instructor ratio is greater than 10/1 or the class size is greater than 20, oral tests are probably not feasible because of time constraints.

- **Environmental Limitations.** If space limitations prohibit testing where other students cannot overhear the examiner, then oral tests should not be used. Written tests should be used instead of oral tests if there is excessive noise involved in oral testing.
4.8. Number and Format of Test Items. If there are a large number of test items or majorities of items are multiple-choice or matching, then written tests should be used because of time and memory constraints.

4.9. Norm-referenced Testing versus Criterion-referenced Testing:

- **Norm-referenced testing** compares the performance of one student with other students.
- **Criterion-referenced testing** allows the instructor to judge the student's performance against a standard established for that task.

4.10. The use of these terms has in the past sparked a great debate among educators involved with testing in technical training. The debate has centered on which of the two is the most appropriate means to interpret scores. For example, two identical tests can be administered to a group of students. One test may be criterion-referenced because the grade was based on a standard criteria (type 45 words a minute with less than three errors) and the grade received by the student is not compared with anything but the successful/unsuccessful completion of the task. On the other hand, norm-referenced testing would take the same results and compare one student with the other students, or groups of students.

**NOTE**

It has been agreed upon that both types of tests are applicable to technical training but the use of the terms criterion and norm-referenced has been de-emphasized. The purpose of testing is to measure a student's achievement of an objective. How these results are interpreted is not the primary goal.

**SUMMARY**

The issues discussed in this chapter include:
(1) Purpose of a Testing Program; (2) Methods of Measuring Objectives; (3) Functions of a Testing Program; (4) Methods of Determining Course Grades and (5) Types of Tests. The next chapter will discuss an overview of test design.
CHAPTER TWO

INTRODUCTION TO TESTING DESIGN
INTRODUCTION

Probably one of the most critical elements in a testing program is the development of tests that measure the objectives. This includes much more than sitting down with references and lesson plans and writing out steps in a performance test or developing multiple-choice test items. This process requires thought and consideration early on in the development/revision project. The developer must begin thinking about testing as the objectives are being written. Testing programs cannot overcome poorly written objectives; therefore, it is important to consider what information will be tested, and how it will be tested as the objectives are being written. This process is called test design.

When a test actually measures the objective(s) to the performance level specified, the test is said to be valid. Thus the purpose of designing a test prior to developing it is to ensure that the test measures student achievement of the objectives and is therefore valid. It is the responsibility of the curriculum developer to ensure that all tests developed during the development/revision project are valid. It is also the responsibility of the course managers and LSO to ensure that tests remain valid during the life of the course.

SECTION ONE - TEST DESIGN

2.1. Test design is the process of determining the physical structure of a test: how the objectives will be measured (performance tests, knowledge tests or practical work), what type of test to use (pretest, progress, comprehensive, etc.), how long the test should be, the learning level of the objectives, number of items per objective, etc. The end result or product of the test design process is a blueprint by which the test, and future versions of the test, will be developed. Test design takes place after the objectives have been written and before any curriculum is developed. Test design applies to both knowledge and performance testing although there is some variation in the two processes. (Refer to Figure 2-1 on page 2-7.
for a flowchart representing an overview of the test design process.) In order to complete the test design, first analyze the learning objectives as they relate to job-specific training requirements.

2.2. Analyzing Objectives for Criticality: This step in the test design process helps determine the criticality of the objectives. Criticality deals with the importance of that objective as it relates to performance of a job-specific requirement. The criticality of the objectives can range from most critical to least critical. In other words, some objectives are more important to actual OJT performance requirements. The results of this analysis are used to assist in determining which of the objectives will be measured by performance tests, which will be measured by knowledge tests, which will be measured by practical work and which will be measured by inference. The following paragraphs provide a general description while Chapters 3 and 4 discuss specific requirements for performance and knowledge test design respectfully.

- The first step in analyzing objectives for criticality is to separate the objectives into two categories, if applicable: Performance Objectives (PO) and Knowledge Objectives (KO). For most courses, the goal is to teach a job-specific skill. With this in mind, the POs are normally more critical than the KOs. There are times when some KOs, when analyzed against other KOs, become critical to the student's ability to perform a follow-on skill. In these instances, KOs may also be considered critical.

- The second step is to analyze each group of objectives to determine that most critical to successful job performance. This analysis must be accomplished by Subject Matter Experts (SMEs). The results are more reliable if there are several SMEs analyzing the objectives. The following methods are listed as examples of how criticality may be determined.

2.3. Rank Ordering of Objectives. This method ranks objectives from most critical to least critical. A cutoff point between
most critical and least critical objectives should be determined. For example, objectives ranked in the top 70% may be considered most critical and would require performance testing. The lower 30% may be considered least critical and may be measured through methods such as practical work or inference. Rank ordering of objectives is easy to use when the course has only a few objectives. Courses with many objectives should use the next method.

2.4. Establishing Categories of Criticality. This method requires SMEs to look at the objectives in categories, such as highly critical, critical, and less critical. There is no ranking involved in this method. Objectives categorized as highly critical, for example, should be measured through performance or knowledge tests while those less critical should be measured through practical work or inference. Regardless of which method is used, SMEs should consider several factors.

- **Frequency of Performance in the Job.** A task frequently performed on the job should be considered critical regardless of difficulty or importance to job performance.
- **Safety to Equipment and Personnel.** Tasks that are considered high risk or dangerous to either personnel or equipment will be considered critical.
- **Importance of the Objective to the Overall Course Mission.** Some objectives that are not as important to the job as are others but are necessary in order to accomplish the course mission and therefore, should be measured by performance or knowledge testing.
- **Importance of the Objective to Performance on the Job.** Objectives that are performed immediately upon graduation should be considered critical.
- **Difficulty of the Objective.** More emphasis may need to be placed on measuring a difficult objective than an objective that is easier to achieve.
- **Attention to Detail.** If the objective requires attention to detail, more emphasis should be placed on its measurement. For example: A course has 51 objectives, 25 performance and 26 knowledge. The POs are analyzed first
to determine which ones to measure through performance tests. The criticality of each objective is considered as it relates to the above categories.

**NOTE**

Whenever possible, the most critical POs should be measured by a performance test. Once it has been determined how to measure the POs, the next step is to consider the KOs.

2.5. Two additional factors should be considered when analyzing KOs.

- **Importance of the Objective as it Relates to Performance.** If the knowledge taught is necessary for the student to perform a skill, it should be considered critical. For example, if it is important for a student to understand transistor theory prior to troubleshooting a particular radar device, then it would be necessary to measure the student's knowledge of transistors through a knowledge test prior to the performance test.

- **Inference (Measurement by Accomplishment of a Higher Objective).** Some objectives may be measured during the measurement of higher-level objectives. For example, if the KO requires the student to state the purpose of a Maintenance Instruction Manual, it might be possible to measure this objective while the student is using the manual in a performance test. This assumes the student must know the purpose of the manual in order to know which manual to go to for the necessary information. To return to the earlier example, there are 26 KOs that must be analyzed for criticality. Each of the above factors are considered when determining how to test. The most critical KOs should be measured through knowledge tests. Those less critical may be measured through quizzes, problem sheets, homework assignments or inference. While there is no formal documentation required for analyzing objectives for criticality, it is a necessary to keep track of how the objectives are measured. This information will be used in
the Testing Plan. A memorandum to file with a list of SMEs conducting the analysis is an example of tracking.

Once the objectives have been analyzed, the process is divided into two paths: develop performance tests and develop knowledge tests. (Refer to Figure 2-1) The accomplishment of this step may occur at anytime but is normally completed after the lesson plans are developed.

**SUMMARY**

The following issues were discussed in this Chapter, (1) Test Design and (2) Analyzing Objectives for Criticality. Performance test development will be discussed in the next chapter.
Analyze Objectives for Criticality

Determine Objectives to be measured by Performance test/practical work.

**PERFORMANCE**

- Develop Job Sheet
- Develop Evaluation Instrument
- Develop Grading Criteria

**KNOWLEDGE**

- Determine Learning level of the objects
- Determine Number of items required to Measure the Objective
- Write items that meet test Design Requirements
- Establish the Minimum Pass Score for the Objectives and the Test.

Administer Test

**Figure 2-1 - ANALYZE OBJECTIVES FOR CRITICALITY**
CHAPTER THREE

PERFORMANCE TEST DEVELOPMENT
INTRODUCTION

3.1. In Chapter 2, Introduction to Test Design, the objectives were analyzed and a determination was made on which objectives require performance testing. The next step is to develop the tests. When this step actually occurs will vary between different development/revision projects. Two factors should be considered in determining when to develop performance tests. The first factor is they should be developed prior to the knowledge tests. How KOs are tested may be contingent on the performance tests. The second factor is to develop the test far enough in advance of the pilot so that validation can occur. Remember the most critical POs should be measured by performance tests unless constraints such as time, equipment, space, etc. prevent it. The following section is an overview of the factors to consider when developing performance tests.

3.2. Factors to Consider when Developing Performance Tests

- **Performance tests** are simulated work experiences where students demonstrate the ability to complete a job. They are used to measure skill objectives in the course. While it may be desirable to measure all skill objectives by performance tests, it may not be practical. This is why it is so important to determine the criticality of the objectives before the first tests are developed. When determining the practicality of measuring all critical objectives by performance tests, several factors should be taken into consideration. These factors are discussed below.

- **Availability of Training Devices.** Sometimes labs may not be equipped with training devices that allow performance of the objective or there may not be an adequate number of training devices available for practice, remediation, or performance retesting. If this occurs, other means of measuring the objective must be established. One possible method is to simulate the performance. On some occasions, the training devices are capable of measuring more than what the objective requires. Caution must be taken to ensure the performance tests are based on course objectives and not on the capability of the training devices or the desires of the curriculum developers.
Space Limitations. When the space available for performance testing is limited, it may not be possible to conduct all the performance tests as required by the objectives. If this occurs, simulation or pencil and paper performance tests may be necessary.

Student Loading. When determining the amount of performance testing, consider the expected maximum student loading and develop tests to accommodate the maximum.

Testing More than One Person. Sometimes performance testing may involve more than one person. This may occur because of limited resources or because of the nature of the task (e.g., fire fighting, damage control). If each student on the team performs different tasks, job sheets should be developed for each task. If each student on the team performs all steps in the task, a job sheet covering the complete task should be developed.

3.3. Measuring Team Performance: When measuring team performance, students should be evaluated according to their individual performance and/or participation in the team. Each student should participate in at least one active position and be evaluated while performing at least some of the steps associated with the objective. Students should not be signed off just for observing or reading a maintenance or operator manual. The determination as to how students should be evaluated during team exercises is based on the objectives. The objective may require students to successfully perform each position on the team or to just serve as a member. The performance test should be developed to measure the objective.

Type of Measurement. The type of measurement is either product, process or a combination of both. Guidelines for determining when to use each type of measurement are discussed later in this chapter.

Type of Grading Criteria Required. The grading criteria may be established as a YES/NO checklist or a rating scale that assigns a numerical value. Guidelines for determining the type of grading criteria to use are discussed later in this chapter.
3.4. Develop Job Sheets: Once it has been determined where performance tests are required, and whether performance testing can be accomplished, the next step is to develop job sheets. A job sheet directs students in the step-by-step performance of a task or job required by the objective. The job sheets developed for performance tests provide a means for students to apply knowledge and skills attained through instruction. Most performance tests are open book in nature and require students to use the technical documentation in completing the task. The job sheet is divided into five sections: Introduction, Equipment, References, Job Steps and Self-Test Items (optional).

- **Introduction.** This section describes the purpose of the performance test, the objectives to be measured, and the benefits/consequences that can be expected for successful/unsuccessful performance.
- **Equipment.** This section lists all the equipment required to complete the performance test. This may be left blank if the objective requires students to select the correct equipment for the job.
- **References.** This section lists the publications required to complete the performance test. This section may be left blank if the objective requires students to select the correct technical manual for performing a particular job.
- **Job Steps.** This section lists procedures for students to follow during the performance test. The amount of detail will vary depending on the objective. For example, if the objective requires students to make decisions and work independently, this section may be less detailed. If the objective requires the student to perform specific procedures and/or requires supervision, then this section may require more detail.
- **Self-Test Items.** This section is optional. It is particularly beneficial if the objective requires students to draw conclusions about the actions taken.

3.5. Types of Measurement:

3.5.1. Performance tests may measure, depending on the objective and the job requirements, a product, process or a combination of both.
Product Measurement. A product is an observable result, (something that can be seen, heard or touched). A solder joint is a product because it should be smooth to the touch. A completed form is a product because it can be seen. Product measurement may be possible when the objective specifies a product. For example, a course objective for a mess management specialist may be to prepare a dinner for a visiting dignitary. Product measurement may also occur when the product can be measured as to the presence or absence of certain characteristics. For example, does it look right, is the texture right, are the wires connected correctly?

Another circumstance for product measurement occurs when the steps of the procedure can be performed in a different order or sequence without affecting the product. For example, SET UP and CONNECT a Simpson 260-5P multi-meter for measuring resistance.

NOTE

one can look at the multi-meter after the procedure is performed to determine if it is correctly connected and set up.)

3.5.2. Process Measurement. Process involves step-by-step procedures required to produce a product or complete a task. Process measurement is appropriate when the product/process are the same thing such as teaching a lesson. Process measurement may also occur when there is a product, but safety, high cost, or other constraints prevent the product from being measured. Process measurement may be used to diagnose reasons for performance failure or when there is a product with critical points in the process which if performed incorrectly, may damage personnel or equipment. Measuring the process should occur when the objective specifies a sequence of steps to be observed or the process does not result in a product. Finally, process measurement may be appropriate when the behavior is the action requiring measurement.
3.5.3. Both Product and Process Measurement. Sometimes you may use both types of measurement in the same performance test. When deciding whether to use product, process or both, measure the product if the objective contains specific standards that the product must meet.

3.5.4. Measure the process if the objective gives specific standards that must be followed during the process. Typical standards include; safety procedures, time standards or the steps must be performed in a certain order. Also measure the process when diagnosis is important (i.e., when/where the errors occur).

3.5.5. If either process or product can be measured, select the easiest to measure. Factors that should be considered: (1) Time or the number of personnel required to do the measurement. (2) Can the product be measured without looking at the process? (3) If errors are made early in the process might they be costly or dangerous?

NOTE

It should be noted that few performance tests are 100% process or product. Most are a combinations measuring steps in a process and a final product.

3.6. Evaluation Instrument: Once the type of measurement has been determined and the job sheets developed, the next step is to develop the evaluation instrument. The evaluation instrument is a formalized checklist or rating scale that is used to guide the instructor in grading the performance test. Both checklists and rating scales may be used for performance tests that measure either process or product. A checklist and/or rating scale should be developed for each task or group of tasks on the job sheet. Guidelines for deciding whether to use checklists or rating scales for both product and process are given in the following paragraphs.
If the characteristic of a product or step in a process is either present/absent, done/not done, and can be measured by checking "yes" or "no," a checklist should be used.

If a product or process can vary in quality from high to low, adequate to inadequate, good to bad, or some other type of scale, a rating scale should be used. It is not necessary that all the characteristics of a product or process use the same type of evaluation instrument. Some may use checklists; some may use rating scales; other may use both. Rating scales can also be used when a step has more than two possible outcomes.

When deciding which evaluation instrument to use, it must be pointed out that it may not make a big difference as to whether a checklist or rating scale is used because almost all rating scales can be turned into checklists, and some checklists can be made into rating scales.

One factor that must be considered concerns how the course is graded. If the course is graded "SAT" or "UNSAT", a checklist would be appropriate. If the course is graded with a numerical grade, a rating scale would be appropriate.

The important issue is to describe the checklist steps/rating scale decisions as precisely as possible. The more precisely the behaviors are described, the more effective the instrument will be.

3.7. Grading Criteria: Once the job sheets and evaluation instruments have been developed, the next step is to develop the grading criteria. The grading criteria is the tool instructors use to ensure that students have achieved the objectives. The grading criteria must be developed in enough detail to ensure an unbiased, objective evaluation of student performance. The performance observed and measured must be consistent with the objectives and their criticality.

3.7.1. To develop grading criteria, job sheets must be analyzed to determine how each step or group of steps will be graded. When deciding how to grade, first determine if there are any steps when performed unsuccessfully, will result in a failure of the performance test. For example, if the safety steps are not performed successfully, it may be necessary to fail the student
on the entire test regardless of the overall grade. This determination **must be** made on the front end, not after the test is graded.

- **Related steps** may then be grouped for grading purposes. For example, instead of grading steps one through five separately, they may be grouped into one graded checkpoint.
- **With the critical steps** identified and grouped for grading purposes, the next step is to determine how to grade each step or group of steps. This will vary depending on the type of evaluation instrument - checklist/rating scale.
- **For the YES/NO checklist**, a detailed description of what constitutes satisfactory and unsatisfactory performance is provided. For product testing, this usually requires a description of the characteristics of a good product. For process testing, this usually requires a description of the correct procedure.
- **For the rating scale**, a detailed description of how the student's numerical grade is determined is provided.
- **For product testing** the following should be provided:
  - A description of how many points each characteristic is worth.
  - How many points are to be deducted for specific errors?
  - How many trials are allowed for each product?
  - Are there any characteristics that if not present, students will fail the test?
- **For process testing** the following should be provided:
  - Description of how much each step/group of steps is worth.
  - How many points are to be deducted for specific errors?
  - How many trials are allowed for each step/group of steps?
  - Are there any steps if not performed correctly, students will fail the test?
Other important factors include:

- Checking operation of equipment after completed assembly.
- Physical testing of the finished job.
- Time required to complete the job.
- Skill in using tools.
- Compliance with required safety precautions.
- Care and use of the equipment.

When any of the above factors are used, a detailed description of what students are expected to do and what happens if they fail to comply with the requirements must be provided.

Oral or written test items may be used on a job sheet to measure the knowledge supporting the performance of a skill. They are normally used at the end of the job sheet to allow students to draw conclusions about the performance. They may also be used at the beginning of the job to test for memory of safety precautions, test equipment, procedural steps, etc.

**NOTE**

When used in conjunction with performance tests, oral or written test items will be written down in the administrator's guide. Oral or written test items should not be used if they interfere with the performance of the task.

3.8. Minimum Passing Grade for Performance Tests

- The next step is to determine the minimum passing grade for performance tests. To accomplish this, SMEs should review the job sheet, evaluation instrument, and grading criteria.
- If a numerical grading system is used, point values that represent minimum acceptable performance, should be assigned to each task on the job sheet or grading criteria. The total point values represent the minimum passing grade for the test.
• If the grading system is SAT/UNSAT, minimum acceptable performance must still be determined. For example, a performance test has seven tasks that are graded SAT/UNSAT. How many of these tasks must be completed for minimum acceptable performance?

• Care must be taken when using SAT/UNSAT grades for performance tests if numerical grades are assigned to the knowledge tests. If this occurs, the student's grade for the course may be based solely on knowledge. This does not indicate a realistic picture of the graduate.

• An explanation of how a grade is determined on performance tests is required in the testing plan.

3.9. Validation of Performance Tests: Validation of performance tests involves several major processes.

• The first process begins with a review of the job sheets and grading criteria for content validity. A test is considered to have content validity if it measures what it is designed to measure. In other words, does the performance test measure the objective(s)?

• The second process begins with administering the test to a group of SMEs. This includes using the grading criteria and the evaluation instrument. Feedback from the SMEs should be gathered and analyzed and changes made as required.

• The third process involves administering the test to a representative sample of students, gathering and analyzing data and making changes as required.

• The final process involves administering the test to the normal classroom, continuing to gather and analyze data and make changes as required.

3.10. Administrator's Guide

The final step is to develop the administrator's guide. The administrator's guide contains the following: Instructions to the Student, Instructions to the Administrator, Evaluation Instrument, Grading Criteria. The administrator's guide may also contain a copy of the job sheet, although it is not a requirement.
• **Instructions to the Student.** This section contains the following:

  - Purpose and description of the test.
  - Explanation of job tasks and what the student is required to do.
  - The level of assistance permitted.
  - How the grade will be determined, including the critical steps which may result in mandatory failure of the test.
  - A list of tools, test equipment, and training material.
  - The allocated time limit and the importance of time to the test grade.
  - Safety precautions.
  - General test regulations.
  - Relationship of the test to the PO being tested and an explanation of the consequences of unsuccessful performance.

• **Instructions to the Administrator.** This section includes the following:

  - A brief description of the task to be performed and a list of tools, test equipment, and training material required. If selecting the proper material is part of the test, it should be so noted in this section. If students are not being evaluated on set-up, provide instructions on how to set-up the equipment.
  - Instructions on special precautions/procedures. If knowledge or oral test items are being used, instructions should be provided to the administrator.
  - Guidance on the actions to be taken in the event that the student does not perform as required. For example, if a critical step is improperly performed, what type of remediation and retesting is necessary?

• **Evaluation Instrument.** Regardless of whether a checklist or rating scale is used, the instrument should contain:

  - A list of steps to be evaluated. This information is taken from the appropriate job sheet. Sometimes it is not possible to evaluate each step separately. If this
occurs, the job sheet will be reviewed and where possible, individual steps will be grouped into like areas and evaluated as one step or checkpoint.

- Each step or group of steps will be numbered in the "Step" column. Also include a brief description of the step in the "Step Description" column. If there are any critical steps that must be completed satisfactorily, they should be indicated.

- Space should be provided for comments or description of errors. Also include space for administrative information as required. For example, name, SS#, class, beginning and ending time, score, etc. Refer to Appendix B for sample Evaluation Instruments.

- Grading Criteria. This is one of the most critical steps in performance test development because it ensures standardized grading between instructors and for students and should contain the following:

  A description of how each step or group of steps is graded.

  If knowledge test items are used in the performance test, list the correct response and indicate how many points will be deducted for an incorrect response. If knowledge test items are used, they should not constitute a major portion of the student's overall grade. In some cases, the knowledge and performance test scores may be kept separately and not combined for an overall grade.

**SUMMARY**

The following issues were discussed in this chapter; (1) Factors to Consider when Developing Performance Tests; (2) Developing Job Sheets; (3) Types of Measurements [process or product]; (4) Grading Criteria; (5) Evaluation Instrument [checklist or rating scale]; (6) Minimum Passing Grade for Performance Tests; (7) Validation of Performance Tests and (8) Administrator's Guide. The next chapter will provide guidelines on how to design a knowledge test.
CHAPTER 4

KNOWLEDGE TEST DESIGN
INTRODUCTION

Knowledge tests are necessary to evaluate the student's ability to recognize, recall, or comprehend facts, procedures, rules, principles, or concepts that are required to perform a skill. Knowledge tests will be developed after the performance tests are written.

4.1. Knowledge test development actually occurs in six different phases:

- Determining the criticality of the KOs
- Deciding what knowledge should be tested and how to test it
- Designing the content and level of difficulty of the test
- Writing test items to satisfy the design
- Determining the minimum passing score and
- Evaluating the test for content validity

NOTE

The first phase was discussed in Chapter 2. Phases two and three are discussed in this chapter. Phase four will be discussed in Chapter 5 while the remaining phases will be discussed in Chapter 6.

4.2. Knowledge Test Design

- As stated earlier, knowledge tests are designed to measure the accomplishment of KOs. It may not be practical to measure all the KOs in a course with knowledge tests. Using the results of the KO analysis, the most critical KOs have been identified. These objectives should be measured by knowledge tests. Less critical objectives may be measured by other means such as quizzes, problem sheets completed in class, homework assignments or inference. (Recall that inference is the method of measuring lower level objectives within higher-level objectives.)
Regardless of the method used to measure the objective, a tracking system should be established to ensure that all objectives, both knowledge and performance are being measured. Refer to Figure 4-1 (page 4-6) for a list of test design elements.

4.3 Learning Level: With the objectives requiring measurement by knowledge tests identified and decisions made on how best to measure those objectives, i.e., progress tests, quizzes, open or closed book, oral or written tests; the next step it determine the learning level of the objectives. There are five learning levels: recognition, recall, comprehension, application, and analysis/evaluation.

• **Recognition.** Recognition is the verbatim identification of specific terms, facts, rules, methods, principles, procedures, objects, etc. that have been presented during training. For example, an objective that requires students to identify parts of a motor by matching their names to a diagram of the motor is a recognition objective.

  • If the objective is written to the recognition level, then all the test items must be written to the recognition level. In a recognition test item, the stem normally requires the student to select the correct answer from two or more alternatives. For simplification purposes, recognition is often abbreviated K1.

• **Recall.** Recall is the verbatim remembering of specific terms, facts, rules, etc. A recall objective requires students to respond from memory instead of selecting from two or more alternatives. Objectives requiring students to list steps in a maintenance procedure or label from memory the parts of a motor, are examples of recall objectives.

  • Recall is a higher learning level than recognition and is often abbreviated K2. If an objective is written to the recall level, some of the test items must be written to the same level in order to measure the objective as specified. Other items may be written to the recognition level but students can not accomplish a recall objective without some recall items. A test item written to a
recall level requires students to remember and respond exactly as taught; therefore, recall is tested with closed book tests.

- **Comprehension.** Comprehension is understanding what was taught rather than memorizing the words. Comprehension objectives normally require students to interpret, explain, translate or summarize information. For example, an objective written to the comprehension level might require students to explain the operation of the motor, not just recognize what it looks like, or label its parts.

  * Comprehension is a higher learning level than recall and is often abbreviated K3. If the objective is written to the comprehension level, there must be an adequate number of comprehensive test items to measure the objective although some items may be written to the lower levels.

- **Application.** Application involves the ability to use acquired knowledge in a situation not specifically demonstrated in class, but job related. Application objectives require students to demonstrate knowledge through mental skill exercises such as determining resistance values from circuit diagrams.

  * Application is a higher learning level than comprehension and is often abbreviated K4. If the objective is written to the application level, there must be an adequate number of application items to measure the objective although some items may be written to the lower levels. The test items must state problems different from the ones used in class to be considered application. If the problem is the same, students may be memorizing the problem and the item becomes a recall item.

- **Analysis/Evaluation.** Analysis involves the understanding of the elements of data and relationships among the data that make meaning of information explicit. Evaluation involves the judgment of value or effectiveness of procedures or solutions based on data, criteria and standards. For example, an item that asks students to select the best approach to meet a stated objective would
require them to know or determine which options would meet the objective (analysis) and which single option would be best (evaluation).

- Analysis/evaluation is a higher learning level than application and is often abbreviated K5. If the objective is written to this level, there must be an adequate number of analysis/evaluation items to measure the objective although some items may be written to a lower level.

4.4. Number Of Items Required to Measure the Objective: After the learning levels have been established, the next step is to determine the number of test items required to measure the objective. Although no formula exists for making this determination, the following guidelines should be considered.

- **Criticality of the Objective.** When both most critical and least critical objectives are measured on the same test, the most critical objective(s) should have more items to ensure the test reflects the critical aspects of the course.

- **Instruction Time Allotted to Present the Material.** If the majority of the time on the course master schedule covers one objective, then the majority of test items should cover that objective. For example, if a progress test covers one week of material and of that week's material, three days was spent on one objective, then the number of items on the test should reflect that time factor.

- **Learning Level of the Objective.** When the learning level is high, such as application or analysis/evaluation, more test items may be needed to ensure understanding. These items must be written no higher than the objective specifies; some items must be written to the learning level specified, while some may be written to the lower levels. If for instance, the objective requires a building block of information, it may be necessary to ask lower level items in order to determine where students are experiencing difficulty.
NOTE

Generally, the more items used, the more accurate the results will be in measuring student competency. For example, a student may be able to guess correctly if an objective is measured by three items, but if ten items are used, it becomes evident that the student understands the material.

4.4.1. There may be times when the objective is written to such detail that only one item can be written. If this occurs, do **not rewrite** the objective but rather group the objectives into content areas. Content areas contain objectives that are similar in content and learning level. Grouping objectives by lessons may also be appropriate. This means that during this step a determination will be made as to how many items are required to measure a content area or a lesson.

4.5. Select Items from the Test Item Bank: The next step is to select the appropriate number of items from the test item bank. If an adequate test item bank already exists, developing items may not be necessary. However, during a development/revision project, the SME cannot proceed until the items are developed.

SUMMARY

The following issues were discussed in this chapter; (1) Knowledge Test Design; (2) Knowledge Learning Levels (recall, recognition, comprehension, application and analysis/evaluation); (3) Number of Items Required to Measure the Objective and (4) Selecting Items from the Test Item Bank. The next chapter will discuss test item construction.
### Cognitive Level of Question

<table>
<thead>
<tr>
<th>WBS#</th>
<th>NLOS</th>
<th>Recall (Fact)</th>
<th>Application (Process)</th>
<th>Synthesis (Principle)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Multiple Choice</td>
<td>True/False</td>
<td>Matching</td>
</tr>
<tr>
<td>1</td>
<td>Define stress IAW OSC doctrine continuum with 100% accuracy</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**FIGURE 4-1 SAMPLE TEST DESIGN FOR A PARTICULAR TEST**
CHAPTER 5

TEST ITEM CONSTRUCTION
INTRODUCTION

After a knowledge test has been designed, the next step is to develop test items that meet the requirements of the test design. Remember when developing test items, use items that measure the important information in the course. Do not measure insignificant or trivial items. An item may be written that meets the criteria listed in the following pages but still be ineffective if it does not measure a valid requirement. In this chapter, five types of test items will be discussed - multiple-choice, true/false, matching, completion or short-answer and essay.

5.1. Multiple-Choice Test Item

- A multiple-choice item is a test item that contains a stem followed by a list of alternatives. The stem contains a statement of the problem and the alternatives contain the correct answer along with several incorrect answers.

- The multiple-choice test item has many advantages and because of these advantages, it is the most frequently used type of test item in technical training. One advantage is it is easy to grade and lends itself to test item analysis. Since it also requires little writing from the student, a greater amount of material can be tested in a relatively short time. It also has the advantage of reducing the effects of guessing. The last advantage is its versatility in that multiple-choice items may measure recognition, comprehension, application, analysis/evaluation type objectives. Refer to Figure 5-1 on page XX.

5.1.1 A disadvantage of multiple-choice test items is they are time consuming to develop and often appear to have more than one correct answer. This last disadvantage can be overcome if the following guidelines are applied when developing multiple-choice test items.

- **Multiple-Choice Test Item Stem Construction Procedures.** A cardinal rule of test item construction is to communicate effectively. For students to understand the question, developers must include all information, conditions,
assumptions, and details to answer the item without referring to the alternatives.

- Consider the following two items. The first item is poorly written question. The second question item has been revised to communicate effectively. See below:

1. A lieutenant wears
   a. A double gold-bar collar device on summer whites.
   b. A uniform appropriate for the occasion.
   c. Two gold sleeve stripes on each sleeve of the winter blue.
   d. An anchor on collar.

2. A Navy lieutenant dressed in a winter blue uniform may be recognized by what uniform designation?
   a. A gold anchor collar device
   b. Two gold sleeve stripes on each sleeve
   c. A double bar collar device
   d. A single gold sleeve stripe on each sleeve

- When the item is referring to a piece of equipment, a subsystem, or a system, the equipment, subsystem or system should be referenced in one of the following two ways.

1. What circuit in the AN/ABC-3Q radar activates the AXO circuit?
   Or

2. (AN/ABC-3Q Radar). What circuit activates the AXO circuit?

- The second method should be used when the placement of the name in stem makes the item grammatically clumsy or difficult to understand.

  - The stem should be phrased positively instead of negatively whenever possible. Negatively phrased stems are often confusing to the student. If a negative is used, it should be either CAPITALIZED or underlined for emphasis.
• Non-essential information should be omitted in the stem.
• A figure number should be referenced within the stem if the test item uses a figure on a separate sheet of paper.

• Designations/symbols for units of measurement for numerical alternatives are included with each alternative. The following example shows how designations/symbols for units of measurement are used properly.

1. How far does the "prohibitive zone" extend from any coast line of the U.S.?
   a. 3 miles
   b. 50 miles
   c. 150 miles
   d. 200 miles

• Include words, phrases, etc., that pertain to all the alternatives in the stem rather than repeat them in each alternative. Consider the following items.

1. An increase in R in a series circuit
   a. results in an increase in Q.
   b. results in an increase in BW.
   c. results in an increase in I.
   d. results in an increase in E.

2. An increase in R in a series circuit results in an increase in:
   a. Q.
   b. BW.
   c. I.
   d. E.

NOTE

Test item number 2 shows the correct way to write this type of test item without repeating in each alternative, the words, phrases, etc., that pertain to all alternatives.
• Connective words such as a/an should be used to ensure that all alternatives are grammatically correct and the single article does not give away the answer. Consider the following example.

1. A one-shot multi-vibrator can be categorized as an ........ multi-vibrator.
   a. blocking
   b. crystal
   c. free running
   d. oscillating

2. A one-shot multi-vibrator can be categorized as a/an multi-vibrator.
   a. blocking
   b. crystal
   c. free running
   d. oscillating

NOTE

Test item number 2 demonstrates the correct method when using connective words such as a/an.

5.1.2. Multiple-Choice Test Item Alternative Construction Procedures. A test item is only as good as its alternatives. Because of this, care must be taken to ensure all incorrect alternatives are plausible but are clearly incorrect. The following guidelines apply when constructing alternatives to multiple-choice test items.

• Normally, multiple-choice test items contain four alternatives, one correct and three incorrect alternatives. In some situations, a credible fourth alternative cannot be written (example: decrease, increase, stays the same). When it becomes illogical to add a fourth alternative, only use three. This approach should be used sparingly in that an item's ability to discriminate between the students who know the answer and the ones who guessed correctly is reduced when the number of alternatives is reduced. In
other words, students have a one in four chance of guessing at the correct answer when there are four alternatives. When the alternatives are decreased to three, the chance is one in three. If the alternatives are decreased further to only two, students have a 50/50 chance of guessing correctly.

- Only one alternative must be clearly correct. It is just as important to justify the reasons an alternative is incorrect as it is to defend the correct alternative. In the example below, "b", "c" and "d" all have the same or similar meaning.

1. When the value of current passing through an inductor changes, the device is considered to be:

   a. reactive.
   b. inactive.
   c. passive.
   d. permissive.

- Interrelated answers such if "a" is true then "b" is false should not be used.
- Vocabulary that is unfamiliar to the students should not be used. Introducing new terms during a test is not recommended.
- Alternatives must be of approximately the same length and complexity, and expressed in similar form. Consider the following items. The first item has the correct answer expressed in a different form. The second item has been revised to correct the error.

1. The two subsystems of 3-M are PMS and:

   b. MDS.
   c. Engineering Operational Procedures.
   d. Engineering Operational Sequencing System.

2. The two subsystems of 3-M are PMS and .......

   a. PQS.
   b. MDS.
Alternatives must all be related, meaningful and not subject to automatic elimination. In the first item the alternatives are not related. The second item provides alternatives that are more credible.

1. Which of the following substances is the hardest?
   a. Cardboard
   b. Glass
   c. Iron
   d. Diamond

2. Which of the following substances is the hardest?
   a. Concrete
   b. Steel
   c. Iron
   d. Diamond

Incorrect alternatives should be based upon a common misconception of students or how students might incorrectly manipulate figures/interpret instructions. Another possible method would be to develop the incorrect alternatives by looking at the correct alternative and determining how it may be made incorrect.

The words always, never, often, normally, etc. should not be used. These terms are referred to as specific determiners or something that gives an unintentional clue to the correct answer. Test-wise students learn quickly that generalizations such as always, all, never are likely to be incorrect alternatives.

All of the above and none of the above should not be used. All of the above is a bad alternative because students only need to find two defensible alternatives to realize that all of the above must be correct. None of the above is a bad alternative because it stresses what is incorrect, not what is correct.
Negative wording in an alternative should not be used. As with negative wording in the stem, negative wording in the alternative may only serve to trick students.

Alternatives which contain numerical values, letters or alpha-numeric characters, should be arranged in ascending or descending order. Consider the following example.

1. Which circuit provides negative feedback for Q12?
   a. C21
   b. L10
   c. R22
   d. R32

With the exception of the above example, the position of the correct alternative must be determined by a random selection process to avoid any patterns which may provide clues to the answers.

5.1.3. Format and Punctuation. It is important to standardize how stems and alternatives are written. There are three different acceptable formats: closed stem, open stem and EXCEPT format. The following guidelines should be used to provide standardization.

5.1.4. Closed Stem test items may be expressed as (1) a question or (2) a statement with the completion position within the body of the stem.

- Question type test items must be complete sentences, punctuated with a question mark at the end of the sentence. Stems should be prepared in question format as much as possible because this type of stem has the advantage of being a clear statement of a problem and the chance for grammatical clues are reduced. The disadvantage is that the question type test item may require more lengthy alternatives.
- Statement type test items should be used sparingly because they are particularly susceptible to grammatical clues and incomplete thoughts.
• When punctuating closed stem test items, the following applies: When the alternatives are complete sentences, each will begin with a capital and end with a period.

1. If the pressure of a gas in a scuba bottle is doubled, what will happen to the temperature and the volume?
   a. Temperature decreases and volume doubles.
   b. Temperature increases and volume remains constant.
   c. Volume increases and temperature remains constant.
   d. Volume decreases and temperature doubles.

• When the alternatives are incomplete phrases or single words, begin the alternative with a capital letter and end it without a punctuation mark.

1. What piece of equipment is used to measure resistance?
   a. Voltmeter
   b. Ohmmeter
   c. Ammeter
   d. Oscilloscope

• When punctuating closed stem statement items, the alternatives should be expressed in such a way as to make the sentence grammatically correct.

1. According to Boyles Law, the volume of air in a scuba diver's lungs will _____ on ascent.
   a. decrease
   b. increase
   c. remain the same
   d. be absorbed

5.1.5. Open stem test items use an incomplete statement with a blank at the end of the test item. The advantage of the open stem test item is it is easier to develop than the closed stem test item.
• When using this type of test item, ensure that each alternative logically completes the sentence and all the alternatives are related to each other. All alternatives must grammatically match the stem and complete the sentence.
• To avoid confusion and to encourage effective communication, only one blank should be contained in the stem and should be placed at the end of the stem.
• The blank should contain some uniform way to indicate that information is missing. Using seven periods to indicate a blank is the method recommended in some development manuals. Another method is to use a blank or an underline to indicate missing information. The important issue is that all blanks be uniform in length and consistent in format throughout the test. The length of the blank should not attempt to indicate the length of the correct response.
• The disadvantage of the open stem test item is there are more chances to provide clues to the correct alternative.
• When punctuating open stem test items begin the alternatives with a lower case letter and end the phrase with a period. If the first word of the alternative is a proper noun, begin it with a capital letter and end the phrase with a period. Consider the example below.

1. Rapture of the deep is defined as _______
   a. DCS.
   b. nitrogen narcisos.
   c. hypoxia.
   d. the bends.

• EXCEPT test items are used when there are three equally correct alternatives. The student selects the incorrect response. The word EXCEPT in the stem must always be capitalized and/or underlined for emphasis.

1. SINS MK MOD 6/7. A specific torquing pattern and associated torque values can be found in the SINS technical manual for all of the following assemblies or components EXCEPT a/an _______
a. azimuth synchro assembly mounted to the stem.
b. velocity meter mounted to the platform.
c. replacement gyroscope mounted to the stable platform.
d. platform stem mounted to the bedplate.

5.2. Common Errors in Multiple Choice Test Item Construction. Errors are underlined.

• Using similar wording in both the stem and only in the correct alternative.

1. What is the purpose of the Mardan maintenance test set?
   a. Monitors the CP operations
   b. Furnishes the power to the Mardan
   c. Functions as a running time meter
   d. Provides static and dynamic testing in the Mardan

• Stating the correct alternative in greater detail that the incorrect alternatives.

1. PWRSS. When all weapons power is removed from the PIP, which of the following statements is true?
   a. All power is lost to the MCC equipment.
   b. The MCC equipment is furnished power from navigation.
   c. Power from the ship control center may be present in MCC since it only goes through the ship JP.
   d. The DCCs have heater power applied.

• Using two or more alternatives with the same meaning.

1. AN/BRN. What is the final step in performing post maintenance checks?
   a. Secure the front panel to the chassis
   b. Return the diagnostic test
   c. Set manual test switch to "OFF"
   d. Make sure the front panel is secure

• Using alternatives that are included in other alternatives.
1. LCHR/HYD. What is the operating time, in seconds, for the pressurization/compensation blow valve to roll from shut to open?

   a. 1 to 3
   b. 1 to 4
   c. 4 to 6
   d. 7 to 10

• Using specific determiners that provide clues to the correct answer.

1. What device is used to sound deep tanks?

   a. Folding rod
   b. Sounding tape
   c. Dipstick
   d. Sight glass

5.3. True/False Test Item

• True/false items are descriptive statements that include all the relevant information and conditions required to correctly answer the question. They are used when there is only one plausible alternative to a problem.

• There are several advantages of the true/false test item. One is it requires less time to develop than do multiple-choice items. It also has the advantage of ease of scoring. Finally, since this type of item is usually short, more material can be measured in a given time period than with any other type of item.

• One of the most obvious disadvantages is it is more susceptible to guessing. The student has a 50 percent chance of answering the item correctly without knowing the correct answer. The true/false item also emphasizes memorization instead of comprehension. Although most do emphasize memorization, it is possible to measure more complex thoughts and ideas with true/false items.

5.4.1. True/False Test Item Construction Procedures. True/false items should test only one central thought or idea at a time.
- The stem must be concise and clear. To ensure this, specific determiners such as always, never, none, all, sometimes, normally, etc., should not be used.
- Stems should also be short. Long stems are harder to read and more difficult to judge true or false.
- A false statement must be consistent with a typical misconception.

(TRUE/FALSE). A Simpson 260 multi-meter may be used to measure impedance.

a. True
b. False

- State each item positively to minimize confusion. The example below may easily confuse the student. The item should be restated positively.

(TRUE/FALSE). Power is not indirectly proportional to current and voltage.

a. True
b. False

- The true/false statement must be clearly true or false. In the example below, the first part of the item is true while the second part is false.

(TRUE/FALSE). Total power dissipation is directly proportional to total current, if \( I_t \) decreases, \( P_t \) must increase.

a. True
b. False

- Tests items should not be lifted verbatim from the curriculum. This encourages instructors to read from the lesson plan and also encourages short term memorization for students. Requiring students to respond to new situations is one way to increase the learning level of the item and avoid memorization.
• True/false items will contain (TRUE/FALSE) at the beginning of each item.
• True/false items may be written to test recognition, comprehension, application, or analysis/evaluation. Refer to Figure 5-1 on page XX.

5.5. Matching Test Item

• Matching test items are well suited to matching terms and definitions. They may also be used to identify points on a schematic or chart. The matching test item is used when it is difficult to develop alternatives, when the items are closely related or when the same alternatives are repeated in several multiple-choice items. It consists of a stem and two columns listed below the stem. The stem contains directions to students on how to match the columns. One column contains the questions and the other contains the answers.
• The matching test item has the advantage of being easy to construct, easy to score and reduces the effects of guessing. A disadvantage is it tends to ask students trivial information.

5.5.1. Matching Test Item Construction Procedures. The stem must specify how students are to match questions with answers, how they are to respond and how often an answer may be used (once, more than once or not at all). The columns should be identified in the stem by title.

• Questions are placed in the left-hand column; answers are placed in the right-hand column. Each item in the right-hand column should consist of single words, numbers, codes, symbols, short phrases, etc.
• To maximize discrimination, each response must be a plausible answer to each question. This means that the answers must all be related to the same area. There should be more or less answers than questions to reduce the possibility of guessing the correct answer through elimination.
• Answers should be arranged in a logical order, i.e., ascending, descending or alphabetically.
To avoid confusion, keep the entire matching test item on one page.

Matching test items are most frequently used for recognition but may also measure comprehension and application. Refer to Figure 5-1.

Match the name of each circuit in Column A with the wiring color in Column B. Enter the letter of a color in Column B in the blank before each item in Column A. There is only one correct answer for each item. No item in Column B will be used more than once.

<table>
<thead>
<tr>
<th>Column A: CIRCUITS</th>
<th>Column B: WIRING COLOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Plates</td>
<td>a. Black</td>
</tr>
<tr>
<td>2. Screen grids</td>
<td>b. Blue</td>
</tr>
<tr>
<td>3. AC power lines</td>
<td>c. Brown</td>
</tr>
<tr>
<td>4. Cathodes</td>
<td>d. Gray</td>
</tr>
<tr>
<td>5. Filaments (off-ground)</td>
<td>e. Green</td>
</tr>
<tr>
<td></td>
<td>f. Orange</td>
</tr>
<tr>
<td></td>
<td>g. Red</td>
</tr>
</tbody>
</table>

Match the types of lamps in Column A with the letters used to represent them in the Classification Chart in Column B. Enter the letter of the lamp in the blank before each item in Column A. There is only one correct answer for each item. No item in Column B will be used more than once.

<table>
<thead>
<tr>
<th>Column A: LAMPS</th>
<th>Column B: Classification Chart of Lamps</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Reflector</td>
</tr>
<tr>
<td>2.</td>
<td>Tubular</td>
</tr>
<tr>
<td>3.</td>
<td>Standard line</td>
</tr>
<tr>
<td>4.</td>
<td>Parabola</td>
</tr>
<tr>
<td>5.</td>
<td>Pear shaped</td>
</tr>
<tr>
<td>6.</td>
<td>Globular</td>
</tr>
<tr>
<td>7.</td>
<td>Candelabra</td>
</tr>
</tbody>
</table>

Match the blocks in the static converter block diagram in Column A with the component in Column B. Write the letter which represents a component in the correct block on the diagram. Components in Column B may be used more than once. There is only one correct component for each block.
### Static Converter

<table>
<thead>
<tr>
<th>Column A: Block Diagram</th>
<th>Column B: Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Filter</td>
<td>a. Filter</td>
</tr>
<tr>
<td>b. Voltage regulator</td>
<td>b. Voltage regulator</td>
</tr>
<tr>
<td>c. Oscillator circuit</td>
<td>c. Oscillator circuit</td>
</tr>
<tr>
<td>d. Frequency sensing</td>
<td>d. Frequency sensing</td>
</tr>
<tr>
<td>circuit</td>
<td>circuit</td>
</tr>
<tr>
<td>e. Control power supplies</td>
<td>e. Control power supplies</td>
</tr>
<tr>
<td>f. Phase control circuit</td>
<td>f. Phase control circuit</td>
</tr>
<tr>
<td>g. Transformer rectifier</td>
<td>g. Transformer rectifier</td>
</tr>
<tr>
<td>h. Scott T transformer</td>
<td>h. Scott T transformer</td>
</tr>
<tr>
<td>i. Dampening circuit</td>
<td>i. Dampening circuit</td>
</tr>
<tr>
<td>j. Teaser invertor</td>
<td>j. Teaser invertor</td>
</tr>
<tr>
<td>k. Main invertor</td>
<td>k. Main invertor</td>
</tr>
</tbody>
</table>

### 5.5.2. Completion or Short-Answer Test Item

- The difference between a completion and a short-answer item is completion asks students to complete the sentence while the short-answer poses a question that can be answered with a word or phrase.

  **Completion:** The square root of 25 is ________

  **Short-answer:** What is the square root of 25?

- Advantages of using this type item include (1) measurement of recall; therefore, the susceptibility to guessing is almost eliminated, (2) relatively easy to construct; and (3) takes less time to read an answer than do multiple-choice items; therefore, a larger number of questions may be used to measure knowledge. This advantage is offset if the items require computation.

- Disadvantages of the completion/short answer test item are it is difficult to score and encourages rote memorization of facts.
• Completion/Short-Answer Test Items Construction Procedures.

Wording must be clear and complete enough to allow the student knowledgeable in the subject to answer it correctly. The following example does not communicate clearly enough for the item to be answered correctly.

1. If the permeability of a core increases ______

• Missing information must be important. This helps avoid measuring trivial information. Consider the following two examples. In the first item, the information omitted is not important. The second item has been revised to stress significant information.

1. An inductor stores electrical energy in the ______ of a magnetic field.

2. An inductor stores electrical energy in the form of a/an ______

• In computational problems, specify the degree of accuracy.

• Use only one blank per item. This eliminates any confusion students may have over what the answer should be. This blank should appear near or at the end of the question. When the blank appears at the beginning of the sentence, students may have to reread the item to discover what they are expected to do. The following is an example of what might happen when several blanks are left in an item.

1. ______ Law states that ______ is ______ proportional to ______ and __________ proportional to _______

• Avoid using statements lifted verbatim from the lesson plan or student material. This encourages memorization rather than understanding. Statements taken out of context, directly from the text, are often ambiguous.

• Avoid the use of specific determiners (always, never, some of the time, etc.).

• Completion and short-answer test items are most frequently used for recall but may also measure comprehension and application. Refer to Figure 5-1, page 50.
5.6. Essay Test Item

- Essay test items differ from the short-answer in degree rather than in kind. The essay item allows greater freedom of response and requires more writing.
- Essay test items have the advantage of giving students freedom to respond within broad limits. Students are allowed to express their ideas with few restraints. Guessing is eliminated because they must provide, rather than select, or identify the correct answer.
- Another advantage is it is practical for testing small numbers of students. Constructing and scoring a lengthy multiple-choice test for ten students probably takes more time than constructing and scoring an essay test. As the numbers of students increase, this advantage decreases.
- Essay test items are also useful for measuring a student's ability to organize, integrate and express ideas.
- Disadvantages of the essay test item include difficulty in scoring objectively. Research has shown there is a great deal of variation in how an essay test is scored. Variation increases when essay items are complex or lengthy.
- Another disadvantage of essay items is they measure only limited aspects of the objective. Because they require more time to write, students must be given fewer items to respond to which is not always a true measure of what students actually know. This disadvantage becomes less serious when the scope of the responses is limited and the number of items increased.
- Grading essay test items is time-consuming for the instructor. Answering the items is time-consuming for students. Students may spend a great deal of time on only one or two essay items, which may limit the item's ability to measure their knowledge.
- Guessing is eliminated with essay items but not bluffing. Students try to fool the instructor by answering something, even if it’s unrelated to the item.
- Most essay items, while claiming to measure higher levels, are really only measuring recall or memorization.
- Essay items also place a premium on writing. Students can read more rapidly than they can write. Time allotted is
often devoted to mechanics of writing, and little time on content. Poor writing skills may also lower the student's score because of poor handwriting or poor grammar.

- Situations that Suggest the Use of Essay Test Items. (1) When an objective specifies writing or recall, an essay item may be necessary. (2) Essay items may also appropriate when the number of students is small. (3) If the test is going to be used only once, an essay test may be more practical than a multiple-choice test.

- Essay Test Item Construction Procedures. Structure the test item by stating clearly and precisely what type of response is required. Asking students to discuss, evaluate or compare, without defining these terms often leads to confusion. Vague or ambiguous items lose the intent.

- Specify limitations by telling students the length of the desired answer and the weight each item will be given when determining the grade. Examples include: how long the answer should be or how long to take when answering the item. Consider the following essay items in the original format and the revised format:

**ORIGINAL**

1. Who was George Washington?

**REVISED**

2. List three accomplishments of George Washington's political career. (1 point each)

- Make each item relatively short and increase the number of items. Long items that require students to tell everything they know about some topic can be vague and difficult to score. With more items, a larger body of knowledge can be measured. Scoring becomes easier because there is less information to read. Consider the following poor examples.

1. Describe in detail your general knowledge. Be objective and specific.
2. Estimate the sociological problems that might accompany the end of the world. Construct an experiment to test your theory.

5.7. Guidelines for Constructing the Grading Criteria. List all the information in the grading criteria that knowledgeable students should be expected to provide to answer the item correctly. If a point system is used, list the weight to be assigned to each item based on information that must be present. If a SAT/UNSAT method is used, the grading criteria must identify the minimum information to be provided in the item. Determine beforehand if partial credit may be awarded.

- Guidelines for Effective Grading. Cover the names on the tests before scoring. This reduces the halo effect or the tendency to allow impressions to influence how the test item is graded.
- Read and evaluate each student's answer to the same item before going on to the next item. It is easier to keep the criteria in mind when measuring a single item than it is to measure all the criteria for all the items.
- Grading the complete test individually has two possible consequences: (1) If the student does well on the first item, there is the tendency to assume equal degree of competence on the remaining items. (2) If the student does very well on the test, the evaluator is likely to rate the next test lower in comparison. The reverse is also true.
- Keep scores of previously read items out of sight when evaluating remaining items. This helps avoid the possibility of letting prior evaluations affect subsequent scores.
- Essay test items may be used to measure recall, comprehension, application and analysis/evaluation. Refer to Figure 5-1 on page 5-21.

5.8. Responsibilities for Constructing Test Items: Normally, it is the curriculum developer's responsibility to write test items. The training is typically provided during the appropriate curriculum developer's course. However, test items may also need to be revised or developed after the course is
implemented. It is important that course managers dealing with testing, SMEs and all instructors be familiar with the material contained in this chapter.

<table>
<thead>
<tr>
<th>TYPE OF TEST ITEM</th>
<th>Multiple-Choice</th>
<th>True/False</th>
<th>Matching</th>
<th>Completion Short-Answer</th>
<th>Essay</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEARNING LEVEL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recognition</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recall</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Comprehension</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Application</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Analysis/Evaluation</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

**FIGURE 5-1 COMPARISON OF TEST ITEMS TO LEARNING LEVEL CHART**

**SUMMARY**

In this chapter the following issues were discussed; (1) Test Item Construction Guidelines for multiple-choice, true/false, matching, completion/short answer and essay items. The next chapter deals with knowledge test development.
CHAPTER 6

KNOWLEDGE TEST DEVELOPMENT
INTRODUCTION

After the test items are written, the next steps are to (1) Verify the content validity of each item; (2) Assemble the test; (3) Verify validity of the test; (4) Establish the minimum passing grade for the test (5) Develop different and alternate tests; and (6) Develop the Administrator's Guide.

6.1. Content Validity of the Test Item

For an item to have content validity it must measure what it is intended to measure to the appropriate learning level and must be technically correct. In order to establish the validity of an item, SMEs should review each item. It is a good idea to have someone other than the test item writer review the item. In addition to the SMEs, it is also recommended that an education or training specialist review the item to detect errors other than those technical in nature. Before an item is placed in the test item bank, it must be reviewed for content validity. To conduct this item review, the following issues should be addressed:

- Is the item technically correct? SMEs should check all the instructional materials to ensure that the item is correctly written.
- Is the correct answer keyed?
- Does the item measure what is required by the objective?
- Does the item measure the knowledge level as indicated by the objective?
- Does the item measure a critical knowledge element of the job? If the answer to this question is no, the item should not be used on a knowledge test.
- Is the item grammatically correct?
- Does the item meet test item construction guidelines?

6.2. Assembling the Test

- With each item valid, the next step is to select items from the test item bank that fit the design. If the test item
bank does not contain an adequate number of items, more items that match the test design may be generated.

- One other option is to reevaluate the test design. This is particularly true if only one, or a few, SMEs were involved in the design process. It is important to get collective ideas of many when preparing this design. This improves the chances of developing a valid test.

- As items are selected, group them according to objectives and place them on the test based on teaching sequence. Grouping by objectives becomes important during the grading process while listing based on teaching sequence helps students perform better. The next step is to review the test for validity.

6.3. Validity of a Knowledge Test—Step One

- Up until now, test items have been reviewed individually. Sometimes, even though the individual items are valid, when placed on a test, the test as a whole may be invalid. Additionally, the test has not actually been administered. Determining the validity of the test is a two step process. First the SMEs should review the test as a whole. The next step is to administer the test to the students. This step occurs after the minimum passing grade for the test has been established. When completing step one, SMEs should consider the following issues:

  - Does the test as a whole reflect the required learning levels of the objectives? This can be determined by reviewing the individual objectives. Recall from Chapter 4, items may be written either to the learning level of the objective or below it. For example, if the objective is written to the comprehension (K3) level and items 1 - 10 measure the objective, is there an adequate number of K3 items to measure the objective? Refer to Figure 6-1, page 6-6.

  - Does the test as a whole reflect the content of the objectives? Once again, by looking at the items measuring a particular objective, this determination can be made more easily.

  - Are there any items on the test that provide answers to other items? This issue may become more important when the tests are computer generated.
6.4. Minimum Passing Grade

- This step has the same purpose as developing grading criteria for a performance test, that being, to determine what minimum acceptable performance is. Since tests are not graded on a percentage, SMEs must establish the minimum passing grade for an objective and a test.

- Several SMEs, normally other than the developers, will review the test to determine which items students must know to barely pass. SMEs must only identify the minimum, not the desired. Critical items should also be identified. A critical test item is one that the students must answer correctly in order to pass, regardless of the students' overall grade on the test. Once the minimum is established, the scores are totaled and an average taken. This average becomes the raw score for that objective or content area. For example, Test A measures three objectives, ten items per objective. Eleven SMEs review the test individually and check those items they feel students must know to barely pass. The 11 scores are totaled and averaged. The resulting number is the minimum passing raw score for the objective(s) or content area(s) and for the test as a whole. Once the raw score is established, the next step is to convert that score to a grade on the scale that was discussed in Chapter 1.

- Not only must the raw score be determined, but all other scores must also have a grade assigned. This is a mathematical process. Some computer grading systems are available to do these computations. (MICROSTASS or ISS).

6.4.1. Translation of the Raw Score on a Knowledge Test to a Grade

**Step One.** Determine the raw score for the minimum acceptable performance. Because the minimum passing is established as barely passing, that number is always equal to 63 (the minimum passing grade on the scale).

**Step Two.** Calculate the grade equivalents for the remaining scores above 63. For example, the raw score for minimum acceptable performance on a test is 17 of 30 items.
Subtract the minimum grade from 100.

\[ 100 - 63 = 37 \]

Subtract the minimum raw score from the total items.

\[ 30 - 17 = 13 \]

Divide the remainder of the grade points by the number of items above the raw score.

\[ 37/13 = 2.84 \]

Add 2.84 to all grades above 63.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Raw Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>63</td>
<td>17</td>
</tr>
<tr>
<td>66</td>
<td>18</td>
</tr>
<tr>
<td>69</td>
<td>19</td>
</tr>
<tr>
<td>72</td>
<td>20</td>
</tr>
<tr>
<td>74</td>
<td>21</td>
</tr>
</tbody>
</table>

**Step Three.** Calculate the grades for scores below 63. This is done by dividing the minimum passing grade by the minimum raw score.

\[ 63/17 = 3.7 \]

Subtract 3.7, starting with 63, for each raw score below 17.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Raw Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>63</td>
<td>17</td>
</tr>
<tr>
<td>59</td>
<td>16</td>
</tr>
<tr>
<td>56</td>
<td>15</td>
</tr>
<tr>
<td>52</td>
<td>14</td>
</tr>
</tbody>
</table>

If the minimum passing grade is higher than 63, the minimum acceptable grade must still be determined first and then the grade translated up to, for example, 69. In this instance, the student may perform at the minimum level but not pass the test because subject matter, level of skill needed in follow-on training, or safety requires the graduate to perform at a higher standard. In the example above, a student would be required to...
answer 19 items correctly to pass the test. This is two items above the minimum of 63. Refer to Figure 6-1 for a sample of a completed test design.

NOTE

In Figure 6-1 that students must answer at least two or three of the highest-level items in order to pass the objective.

<table>
<thead>
<tr>
<th>OBJECTIVE</th>
<th>CRITICAL</th>
<th>TOPIC</th>
<th># OF ITEMS</th>
<th>ITEM K LEVEL</th>
<th>RAW SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.4 (K3)</td>
<td>YES</td>
<td>1-1-1</td>
<td>1 - 10</td>
<td>3 (K1)</td>
<td>6 OF 10</td>
</tr>
<tr>
<td>1.5 (K1)</td>
<td>YES</td>
<td>1-1-1</td>
<td>11 - 15</td>
<td>K1</td>
<td>3 OF 5</td>
</tr>
<tr>
<td>1.6 (K4)</td>
<td>YES</td>
<td>1-2-1</td>
<td>16 - 30</td>
<td>5 (K3)</td>
<td>8 OF 15</td>
</tr>
</tbody>
</table>

**FIGURE 6-1 COMPLETED TEST DESIGN SUMMARY**

6.5. Validity of a Knowledge Test - Step Two

- With the minimum passing score established and after the SME review for validity, the next step is to administer the test to a group of students. They should provide feedback in the following areas:
  - Were the items clear and easy to understand?
  - Was the test reflective of the material taught in class?
  - Was the test too easy?
  - Was the test too difficult?

- Using the feedback, corrections should be made to the test before administering it to the class for grading purposes. If there is a great deal of negative feedback, it may be necessary to administer the test once again to a sample group of students. The process of validation does not stop here but will continue with the test items analysis program covered in Chapter 8.
• After a test and test items have been validated, it may be necessary to generate new items. To accomplish this, ensure that each item has content validity as discussed on page 50 prior to being administered to the students. A good method of administering new test items is too systematically include five to ten new items with each old version of the test.

• Item analysis is collected through enough administrations of the test item to provide sufficient confidence to include the item in the test item bank. A considerable bank of new items of proven value can be developed in this manner in a relatively short time.

6.6. Different or Alternate Test Versions

The test design is important when constructing additional versions of a test. Each version of a test must measure the same objectives and to the same level, in order to be considered to be valid. The test design in Figure 6-1 helps to ensure that this is accomplished.

A different version of a test is used for retesting purposes. To be considered different, it must contain no test items from the original test but must cover the same material at an equal degree of difficulty.

Alternate versions of a test are used when more than one class is on board at a time. Alternate versions may use some of the same test items as the original test. The items may be switched around provided they are still grouped by objectives and according to teaching sequence.

The minimum number of required versions is the original and a different version for retesting. The number of alternate versions of a test will depend on the number of classes convening each year.

6.7. Administrator's Guide (AG)

• There are two components to the administrator's guide, the Test Instructions for the Administrator and the Master Test Booklet.
Test Instructions contain a list of directions that explain to the administrator how to prepare the test area, special instructions for students and the time limit allowed for testing. Also included are instructions for the administrator at the completion of testing on how to secure the test area, how to review, evaluate, or critique the test and record the test results.

Master Test Booklet contains test items and a test answer key. It is compiled at the completion of test item development.

**SUMMARY**

This chapter discussed the following issues; (1) Content Validity of a Test Item; (2) Assembling a Knowledge Test; (3) Validity of a Test (4) Establishing the Minimum Passing Grade; (5) Translating a Raw Score to a Grade; (6) Different or Alternate Test Versions and (7) Administrator's Guide. The next chapter will discuss how to manage a testing program once the tests and test items are considered valid.
CHAPTER 7

MANAGING A TESTING PROGRAM
INTRODUCTION

After the testing materials have been developed and validated, they must be implemented. In addition, there are several other elements of a testing program that must be considered. These elements include maintaining the test item bank, ensuring the security of the testing material, administering a test, grading a test, reviewing missed test items, remediating failed objectives and retesting those objectives as required. While the curriculum developer must be aware of these requirements during the development/revision process, it is the course manager's responsibility to ensure the effective conduct of each element. In this chapter, each of the above elements will be discussed as to requirements and possible courses of action. At the conclusion of this chapter, there is a discussion of who is responsible for ensuring the different functions of a testing program are completed correctly, i.e., CCA, CCMM, participating sites, all activities and testing officers.

7.1. Test Item Bank

- The master test item bank contains all the knowledge test items approved for use in the course. The original test item bank is prepared and validated during the development/revision process. The items in the test item bank are used to prepare different and alternate versions of a test. The number of items contained in the test item bank will be based on the test design and number of different or alternate tests required. The test item bank may have items added, deleted, or revised after the material has been validated.

- Contents of the test item banks should be reviewed during the formal course review. Test item banks may be maintained in the form of test item cards, (refer to Appendix C for a sample card) copies of versions of a test or computer-stored test items. Each item in the test bank should contain, at a minimum, the following:

  - Number of the Objective the Test Item Measures. This is a cross reference used to ensure the test items measure the objectives as required.
The Learning Level of the Test Item. This is necessary to ensure that each item measures the objective to the appropriate learning level.

Location of Supporting Material in the Curriculum. Referencing the lesson, technical manual and/or student/trainee guide is necessary to indicate where the material being measured can be found.

Test Item Analysis Data Information. Keeping this information is necessary to analyze trends that may develop in certain items and tests. This includes the difficulty index, discrimination index and effectiveness of alternatives index, as appropriate. For more information on test item analysis, refer to Chapter 8, Test and Test Item Analysis.

The Designation or Number of the Test(s) and Item Number on which the Item is Located. Since it is possible to use a test item on more than one test, it is necessary to track the test(s) on which it is used. For example, a particular item may be used on Version A, item 5 and on Version B, item 3.

7.2. Test Security

- Test materials include test item banks; copies of the tests; scoring keys; computers containing testing materials and any diagram, formula sheet, etc., used by the student when taking a test. Test materials must be controlled in a way that provides for the maximum security of the information. The following guidelines apply to test material security.
- Test materials should be stored in a locked container in an area accessible to staff personnel only. When test materials are removed, a check-in and check-out system should be established so that an accurate, ongoing inventory system of all tests can be maintained.
- Test materials maintained in word processing centers on tapes or disks and those in draft stages will be secured in the same manner as finalized tests. A computer having test items stored on a hard disk drive will be in an area accessible to staff personnel only. Tests are normally unclassified but are to be handled in an accountable
manner. If the test contains classified materials, it will be handled as per the applicable security classification.

- Performance tests and materials should be controlled only when they contain information that could cause a test compromise. A duplicate set of testing materials should be maintained in a separate location in the same manner as listed above.

7.3. Test Administration

- Tests are administered as required by the Test Administrator's Guide. This guide was prepared during the development/revision process. An additional factor to consider is who is responsible for administering the test. There are two different approaches: one is to have the instructor administer the test and the other is to have the test administered by someone other than the instructor, i.e., subject matter expert, testing officer, etc. Both approaches have advantages and disadvantages.

- Most courses use the instructor to administer the test since this is seen as an extension of the instructor's duties. The instructor is more familiar with the material and can better answer the questions students might have. One disadvantage is the instructor may be needed to teach other classes or prepare to teach other lessons. One other possible disadvantage is the instructor, after administering the test, may become so familiar with the test that the emphasis in class is on the test and not on the objective(s).

- To free the instructor for other duties or to overcome the tendency to teach to a test, some course managers use personnel other than instructors to administer the test. If the administrator is familiar with the course material and can answer questions students might have concerning the test, this may be an effective alternative. If instructors are not used, it is important to involve the them in the grading and review of the material in order to provide the instructors with feedback on student progress.

- Tests should be administered in such a way as to prevent test compromise. These guidelines will be outlined in the AG. If test compromise does occur, the test compromised will not be used again until all classes onboard at the
time of the test compromise have graduated. In addition, each activity will have in place appropriate action for the instructor to take if students are caught cheating. This situation is a discipline problem; not an academic one and should be handled as such.

7.4. Grading the Test

- There are two approaches to grading tests: manually or by use of a computer. Regardless of the method used, the important issue is that tests be graded as soon as possible in order to provide timely feedback to the student.
- Of the two methods, computer grading is the most efficient. There are several different types of programs that are being used to grade tests. For example, Instructional Support System (ISS) is a system that provides data on student tracking, test grading and evaluation, student and course management and test analysis and interfaces with the NITRAS system. For courses on ISS, test answer sheets are scanned and the results provided immediately to the instructor. Two other systems used to grade tests and provide immediate results are the Versatile Training System (VTS) and MicroStass.

7.5. Test Review

- As stated earlier, written guidelines for administrating both performance and knowledge tests will be provided in the form of administrator's guides. During the review of the test, precautions should be taken to minimize the possibility of test compromise. After the test has been given and graded, the instructor or test administrator should review the test with the students. This review process is a key factor in producing a quality graduate. By reviewing the missed items, students are made aware of what was missed and why. When inadequate reviews are conducted, students may graduate from training with incorrect information. Inadequate reviews are also demotivating to students in that feedback on their progress is not provided.
- After the test is graded, a general review of the test should be conducted with the entire class. This is normally
accomplished by reviewing the most frequently missed test items with the class as a whole. When only one or two students miss an item, this item may be reviewed in class or individually depending on the situation and time available. The important issue is that every item a student misses should be reviewed with the student in order to provide the correct information.

- In some cases, the best method of reviewing a test is to return the test to the students and review all missed items. But this may not be possible due to the nature of the training, the number of classes on board and the possibility of test compromise. For example, an "A" school may have a new class convening each week with all students staying in the same barracks. If the tests are returned, it won't take long for the answers to be out and therefore the test becomes compromised. If a class convenes only three times a year, returning the test for review may not cause a compromise.

- In cases when the test is not returned to the students, precautions must still be taken to minimize the possibility of test compromise. The following are examples of methods to prevent test compromise:

  - Review the missed test items without discussing the items or the answers verbatim. This requires the instructor or administrator to paraphrase the item and responses.
  - Use computer-generated testing. When new tests are generated each time, the test items and answers may be reviewed verbatim. This may not be practical for courses with large student input due to the volume of printed material required.
  - Use several alternate versions of a test. Alternate versions of a test will follow the original test design. Using several versions of a test requires the course to have a larger test item bank. The rule of thumb for determining if an adequate number of test versions is available is to have enough versions to prevent two classes that are on board at the same time from being administered the same test version.
7.6. Remediation Programs

- A remediation program is a program designed to aid students in achieving the objectives of the course by providing additional instructional study time. Remediation is necessary because not all students can accomplish the objectives or understand the material in the normal classroom time. Remediation programs are normally classified as mandatory and voluntary. Procedures and factors to consider when establishing remediation programs are discussed in the following paragraphs.

- Remediation is not a form of discipline. When students placed in remediation are made to feel like failures or see remediation as a form of punishment, remediation becomes ineffective. Remediation should be presented with a positive manner. Instructors should approach remediation as a means to provide additional help for those who need it.

  - For example, if a student is sleeping in class or is not turning in homework, the student should be counseled first to determine why the behavior is occurring. Remediation may then be assigned as a result of the counseling session, not as a means to correct the behavior but to help the student study material missed. Therefore, remediation can be used to motivate and assist the student in the learning process.

- Because remediation is such an important issue in student success, instructors, trained and certified in the subject matter, should be available to the students during remediation.

- Remediation is affected by the criticality of the objectives. Recall from Chapter 2, Introduction to Test Design, that critical objectives are objectives required for job performance. Depending of the criticality of the objective not successfully completed, the type of remediation method selected may vary.
7.7. Mandatory Remediation

- Mandatory remediation is remediation assigned to the student. The student has not volunteered and is required to attend this training which takes place outside the normal training day. A student may be assigned mandatory remediation in the following situations:

  ▪ **Recommended by an instructor as a result of a performance counseling session.** This occurs if the instructor feels the student may be lagging behind or needs some clarification of material just taught. The instructor makes the recommendation to the course supervisor who is responsible for deciding what action to take.

  ▪ **Result of an Academic Review Board (ARB) action.** Care must be taken to keep the concept of remediation positive. Since ARBs are convened for academic reasons only, discipline should not be an issue.

  ▪ **Result of poor performance on tests, homework, etc.**

  ▪ **When the minimum passing score on a progress or within course comprehensive test is not achieved.**

  ▪ **When an objective is not achieved.**

- While mandatory remediation may be assigned for any of the above reasons, different approaches to conducting remediation may be used. These approaches include:

  ▪ **Time spent in remediation**

  ▪ **The structure of the remediation environment, which includes how much the instructor is involved and the location of the remediation**

  ▪ **Type of failure.**

- **Time Spent in Remediation.** Depending on the type of failure and the level of student performance, students may spend anywhere from one hour to several hours in remediation. The amount of time required should be based on the individual needs of the student. One factor to be considered is the introduction of new material while the student is remediating on the old material. Because this is not conducive to an effective learning environment,
remediation normally is conducted in the same day as the
test failure with retesting following that night or the
next day.

- **Structure of the Remediation Environment.** The structure of
  the remediation environment may be formal/structured or
  informal/unstructured.

  - **Formal/structured remediation** environment refers to
    remediation that is conducted according to pre-
established, written guidelines for the student on
specific areas of study. Quizzes may be administered
during remediation and instructors should evaluate
student performance. Grades on the quizzes may be
recorded in the student's record. A formal/structured
remediation program requires direct supervision and
active involvement by the remedial instructors.

  - **Informal/unstructured remediation** environment places
    more responsibility on the student as to what to study
and where to study. The remedial instructors will be
less involved with this type of remediation.

- **Type of Failure.** Type of failure refers to whether the
  student failed a complete test or failed the objective. It
also refers to the degree of the failure.

  - If a **test is failed**, students may receive remediation
    on the entire test or the part of the test failed.
    Remediation should be formal/structured if a student
fails a test.

  - If the **test is passed, but an objective is failed**, the
    remediation provided may take one of the following
approaches depending on the criticality of the
objective and the degree of the failure.

  - **If the objective failed is a critical objective,**
    remediation may need to be formal/structured.

  - **If the objective failed is not critical,** remediation
    may be informal/ unstructured.

  - Regardless of the criticality, when the student clearly
    does not understand the objective, remediation may need to
be formal/structured. For example, there are ten items on
a test to measure an objective. The standard for the
objective is 6 of 10 items correct. The student passes the test but only answers 1 item correctly. The remediation for this student should be formal/structured. If another student passes the test but answers 5 of 10 items correctly, remediation conducted one-on-one by the instructor may be appropriate. The student may also be allowed to complete some additional assignments individually in an informal/unstructured environment.

7.8. Voluntary Remediation: Voluntary remediation programs provide assistance for the students who seek additional help on their own. Students should be encouraged to ask for assistance anytime they are confused about the material. If the student volunteers for remediation, it may be necessary to separate the voluntary group from the mandatory group. This is true because students in voluntary remediation may take up so much of the instructor's time that students with more severe problems do not receive adequate assistance. The important issue is to provide whatever assistance the student needs to understand the material.

7.9. Techniques to Use During Remediation: Because students and situations are unique, instructors must be creative in how they provide remediation. The following are examples of different techniques that may be used during remediation.

- **Written Self-study Remediation Packages.** Normally contain additional problems, readings or questions the student answers during an after hours program. The after hours program may be conducted as either a mandatory or voluntary remediation program. Depending on the needs of the student, this method may be used in either an informal/unstructured environment or a formal/structured one. This type of remedial material should be developed for areas that have historically exhibited a high failure rate. For example, if students normally have difficulty with transistor theory, additional remedial materials should be developed to provide specific areas of study for the students.

- **Mini-lectures Delivered to a Small Group of Students.** These lectures are prepared for the areas of the course with high failure rates. They are not a re-teach of the lesson but
rather short lessons on very specific subject matter presented by instructors certified to teach the material.

- **Videotaped Lessons.** Lessons may be videotaped for use by students during remediation. This does not mean that an entire course can be placed on videotape. Lessons that are videotaped should be those that historically cause the students difficulty. When this method is used, the tapes should be cataloged so students can find the specific area needed for remediation.

- **Quiet Study as a Non-structured Type of Remediation.** This is best suited for students with good study habits who have little difficulty in attaining the objectives. Normally this type of student is capable of self-directed study and will need little help from the instructor.

- **Remediation for a Student Who’s Unit/Course Average Falls below a Predetermined Grade.** This technique helps to identify students with potential problems before they experience a failure. For example, a course may have a minimum passing grade of 63. Historically, the student who does not maintain an average of 70 has difficulty in the course. If a student's average falls below 70, it may be appropriate to place the student on remediation. The environment should be informal/unstructured and remediation should continue until the student's average exceeds 70.

- **Remediation for a Student Whose Test Grade Falls Below a Predetermined Grade.** This technique has the same purpose as the previous one but differs in that it is used only when an individual's test grade does not fall within the predetermined grade. For example, the minimum passing grade for a test is 70. Students who do not achieve a 75 or higher on that particular test are assigned remediation. The remediation should be informal/unstructured. If the student scores 75 or higher on the next test, the requirement should be removed.

- **Computer Assisted Remediation.** Where available, computers may be used to provide additional instruction and/or practice. Computers do not replace an instructor in the remediation process. Instructors must be available to make decisions on what materials should be used for remediation and to clarify/augment the computer delivered remediation.
7.10. Retesting Procedures: In addition to the remediation policies, retesting procedures must also be established. Normally, these procedures are contained as a part of the remediation program. As with remediation, retesting procedures are also affected by criticality of the objectives and degree of failure.

7.10.1. Formal Retesting. Formal retesting means the student is given a written or performance retest to complete. Formal retesting should occur when: (1) A test is failed, (2) critical objective is failed, and (3) it is clear the student does not understand the objective.

- When a test is failed, the retest may cover the portion of the test failed or the entire test. This decision should be based on the degree of the test failure and the student's performance on the objectives. For example, if the minimum passing grade is 70 and the student made 68, retesting on the entire test may not be necessary provided the student passed some of the objectives. If another student makes 68 but failed all the objectives on the test, a retest on the entire test would be appropriate.
- If the student passes the material retested, the grade assigned to the retest will be the minimum passing grade for the test. For example, if a student scores 90 on a retest, and minimum passing grade is 63, then the grade assigned for the record would be 63.
- When a test is passed but the student fails a critical objective, the student should be retested ONLY on the objective failed. Performance tests may provide an exception to this rule. If the performance cannot be measured by retesting only the failed objectives, a complete retest may be administered.
- When the test is passed but an objective is failed, regardless of criticality, to the degree that it is clear the student does NOT understand the objective, the student should be retested on the objective failed. For example, the student scores an 85 on a test but fails to answer any items on one of the objectives. In this case, it is clear the student does not understand the material therefore a formal retest on that objective should be administered.
• If the student passes a test but is administered a retest due to objective failure, the grade assigned for record purposes will be the original test grade.

7.10.2. Informal or Oral Retesting.

• Formal retesting may not always be necessary. Informal or oral retesting may be used in the following instances:
  ▪ When the student passes the test but fails a critical objective by only one or two test items.
  ▪ When the student passes the test but fails a less critical objective.

• Retesting will occur as soon as possible after remediation. Prolonging the completion of remediation and retesting may cause the student unnecessary difficulties with the new lesson material.

• Anytime an objective or a test is failed, students will receive remediation and will be retested either formally or informally, except as noted below. The approach and methods of remediation and retesting should be based on the situation.

• Remediation and retesting is not mandatory if the test falls on the last day of training and it is not possible to remediate and retest the student. The student should be referred to an ARB or similar review board. If the board determines that the student has failed to achieve the course objectives, the student may be considered an academic attrite. If the board determines that the student has passed the course objectives, the student may be considered a graduate. This requirement exists for all types of courses not just "A" and "G" courses.

• As per the NADEVTRA 135B, all training activities will develop a remediation program to include procedures for voluntary remediation, mandatory remediation and retesting. All remediation and retesting procedures will be described in the testing plan for the course.
7.11. Testing Program Responsibilities: The responsibilities to ensure a quality testing program are shared by several different levels within the organization. The following is a list of responsibilities assigned to each level:

- Course Curriculum Authority (CCA)s are responsible for resolving any differences between the Course Curriculum Model Managers (CCMM) and the participating activity.
- Training Managers are responsible for monitoring the testing programs at the training activities. Learning Sites (LS)/Detachment (DET) Learning Standards Officer (LSO)s shall provide IS training as required.
- The LC is responsible for the following:
  - Oversee development of the testing plan.
  - Developing, validating, and verifying the initial test item banks (both knowledge and performance).
  - Developing test AGs and grading rubrics.
  - Maintain the Master Test item bank.
  - Reviewing test items submitted by participating activities/LSs for possible inclusion in the master bank.
  - Maintain testing data for test item analysis.
  - Provide the participating activities/LSs with the testing plan and master copies of the test item banks, scoring keys, and test AGs.
  - Provide the participating activities/LSs with updated versions of testing program materials, as required.

- Participating Activities are responsible for the following:
  - Provide comments on the testing plan to the CCMM.
  - Provide timely feedback to the CCMM on testing problems.
  - Submitting test Items to the CCMM for review and approval.
  - Revising/updating the test item bank as directed.
  - Maintain test analysis data.

- All Activities are responsible for the following:
  - Appoint Testing Officer(s).
• Prepare testing materials.
• Administer tests.
• Grading of tests.
• Security of all test materials.
• Coordinating and managing the revisions to the tests.
• Conduct analysis to include test item, knowledge test, and performance test analysis.
• Providing feedback on the analysis results to the course supervisor for possible changes to the curriculum or instructional program.
• Coordinating the IS training needs with LS/DET LSO.
• Providing input to the learning center LSO via the LS/DET/Participating Activity for the quarterly training quality indicator summary.

SUMMARY

In this chapter, the following issues were discussed: (1) Test Item Banks; (2) Test Security; (3) Test Administration; (4) Grading the Test(s); (5) Test Review; (6) Remediation Programs; (7) Mandatory Remediation; (8) Voluntary Remediation; (9) Techniques to Use During Remediation; (10) Retesting Programs and (11) Testing Program Responsibilities. The next chapter will discuss test and test item analysis.
CHAPTER 8

TEST AND TEST ITEM ANALYSIS
INTRODUCTION

In Chapters 1 through 7, two important aspects of testing were discussed. The first provided guidance on how to design and develop tests that measure the objectives. The second provided guidance on how to organize and manage a testing program that ensures all graduates possess the skills to do the job and that those who need additional help, or have not met the requirements, are identified. The final step is to measure the effectiveness of the tests and testing program. This is accomplished through test item analysis, test analysis and a review of the Training Quality Indicators (TQI).

PURPOSE

Analyzing test and test items is a process that helps evaluate the effectiveness of the test and test item. The results of the analysis provide feedback on the test item and the test as a whole. For example, is the test too difficult? Is the test too easy? Are the items ambiguous or miss keyed? Is the test actually measuring what it is intended to measure? Are there any objectives students are having more difficulty with than others? Do some tests have higher failure rates than others? Do some versions of a test appear easier than other versions? The results of the analysis are used to improve the tests, improve instruction, and/or improve the testing program.

8.1. Techniques for Test Item Analysis: The first concept for discussion is test item analysis. Test item analysis looks at the individual items to determine the difficulty of the item and how effective the alternatives are in a multiple choice test item. The following techniques are used to assist in test item analysis: difficulty index (p) and the effectiveness of the alternatives for multiple-choice items. Each of these techniques will be discussed in the following paragraph.

- Item Difficulty (p) calculates the difficulty of the test item, how easy or how hard it is. Expressed in another way, it compares the number of students answering the item correctly with the total number of students answering the item. The acceptable range of difficulty for technical training is .50 to .90.
To calculate the difficulty index, using the complete sample, count the total number of correct answers (Nc) and divide by the total number of students responding to the item (N). The formula \( p = \frac{N_c}{N} \) results in a proportion or decimal that becomes the index of item difficulty. The larger the index, the more students answered it correctly. If everyone answers the item correctly, the index would be 1.00. If no one answered it correctly, the index would be 0.00.

The following example indicates that of the 150 answer sheets, 100 students answered the item correctly. Based on the acceptable difficulty range, this item would be considered acceptable. The .66 means that 66% of the students who answered the item answered it correctly.

\[
P = \frac{100}{150} = .66
\]

Just because an item falls outside the acceptable difficulty range, it does not automatically make that item a poor item. Sometimes, for instance, a difficulty index of 1.00 may be desirable. This normally occurs in an area such as safety where the goal is for everyone to answer the item correctly.

Effectiveness of the alternatives as an index is used for multiple-choice test items only. The multiple-choice test item is only as good as its alternatives. If the incorrect alternatives are illogical, not plausible or absurd, students may be able to select the correct response without knowing the material. This index calculates the number of students selecting each alternative within the high and low groups.

To calculate the effectiveness of the alternatives, sort the answer sheets from highest to lowest; select the answer sheets for the highest and lowest 27% of the students. Count the number of students in each group that selected each alternative.
For example:

<table>
<thead>
<tr>
<th>ITEM 1</th>
<th>(a)</th>
<th>(*b)</th>
<th>(c)</th>
<th>(d)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>High 27%</td>
<td>2</td>
<td>15</td>
<td>17</td>
<td>7</td>
<td>41</td>
</tr>
<tr>
<td>Low 27%</td>
<td>1</td>
<td>12</td>
<td>15</td>
<td>13</td>
<td>41</td>
</tr>
</tbody>
</table>

- Alternative "A" may need to be improved. It is ineffective as an alternative since it was selected by only 3 of 82 students. Alternative "C" is deceiving more students in the high group than the low group; therefore, it discriminates in a negative direction. This item should be reviewed for an improperly keyed answer guide, more than one correct answer, or ambiguity.

8.2. Procedures for Analyzing Test Items

- Test item analysis procedures will vary between courses due to the number of students involved. The following general guidelines are provided. Analyses are conducted from student answer sheets. The recommended sample size is 100. Smaller sample sizes may be used, but the data is not as reliable as with the larger numbers. When 100 answer sheets have been collected, an analysis of the two indexes should be conducted.

- When recording the results of test item analysis, it is important to indicate the dates the items are analyzed and to keep track of the performance of each test item over time. This information may be maintained manually on the test item bank or automatically with some computer programs. This historical data is needed to study trends.

  - For example, if the difficulty index of a test item suddenly changes, testing personnel should investigate possible causes for the change. If a difficult item is now very easy, it may have been compromised. If an easy item suddenly becomes very difficult, this may mean instructors are not teaching effectively or the quality of the student has changed.

- In either case, historical data is required to make decisions about test items over time. The frequency with which an analysis is conducted will vary depending on the
circumstances. While 100 answer sheets is the recommended number to use for analysis, this may not always be possible or practical.

- For example, if a course has a large student input, conducting an analysis for every 100-answer sheets may be too time consuming. If this occurs, testing personnel may be able to conduct a monthly analysis until the items are considered statistically valid. This means that the item falls within the acceptable range(s) and the results remain constant. Once statistically valid, the analysis can be conducted on a quarterly basis.

- Some courses may have a very small student input, and it may take several years to collect 100 answer sheets. For courses with small inputs, the entire sample may be used to calculate the effectiveness of the alternatives. These courses may also use the 50% missed rule. Each test item that is missed by 50% of the students will be reviewed for possible problem areas. If 100 answer sheets can be accumulated in a year's time, then an analysis using all three indexes should be conducted.

- There are times when 100 answer sheets cannot be collected in a year. Requests for a waiver from this policy are explained in the testing plan. If a complete analysis is not required, the 50% missed rule will apply.

- The important thing is not so much that an analysis be conducted every time 100 answer sheets are received, but that an analysis is conducted and the results are used to improve the instruction.

- When test item analysis is being conducted for multi-sited courses, both the participating site and the CCMM are responsible for test item analysis. The participating site is responsible for conducting analysis as described above. They are also responsible for forwarding raw data and the results of the analysis to the CCMM. The CCMM will take the raw data and conduct a complete analysis if 100 answer sheets can be collected from all sites in a year's time. If 100 answer sheets cannot be collected from all sites, the CCMM is responsible for conducting a difficulty index
and an effectiveness of the alternatives using the complete sample.

- When analyzing a test item it is important to record the date an item is changed or the date the instructional materials are changed. Each time an item or material is changed, the analysis must begin again. When this is done, it is possible to compare the performance of the test item before and after the change.

- After the test items are analyzed, the next step is to make decisions based on the data. There are three possible areas of review: the test item, the curriculum, and the instructor.

- First, determine which items do not fall into the acceptable indexes discussed earlier. To analyze the item review it by asking questions such as:
  - Is the answer miss-keyed?
  - Is there only one correct answer?
  - Is the question clear to the student?

- If the test item is determined to be sound, do not change it; rather proceed to the next step. The next step then is to review the instructional material by asking questions such as:
  - Is the information correct?
  - Does the material in the student guide support the information in the lesson plan?
  - Does the information in the technical manual; support the lesson plan material in the lesson plan?

- If the instructional material is correct, then the next step is to evaluate the classroom instruction by asking questions such as:
  - Was the material taught correctly?
  - Did the student receive adequate practice prior to testing?
  - Was there adequate time allowed for review and summary of the material?
  - How effective was the instructor in the delivery?
Can the poor performance of the test item be tracked to a specific instructor?

Once all the information has been reviewed, several possible actions may occur. The test, instructional materials, and/or master schedule may require a change. If the instructor appears to be the problem, this may be corrected through in-service training, either technical or technique in nature. The final possible action is to make no change until further data is collected. It must be pointed out that the data from the test item analysis is to be used as a guideline and does not mean changes have to be made. It does mean that the data will continue to be evaluated for possible future action.

8.3. Analyzing Essay Test Items: If essay type test items are used to measure student performance, the requirement to conduct test item analysis still exists. If computers are used the answers must be placed on a scan able answer sheet. In order to analyze essay test items the grading criteria must first be prepared. This grading criterion identifies how an instructor grades an answer. Without the grading criteria, test item analysis cannot be conducted on essay items. Some grading criteria may provide partial credit for an answer. If the essay item has several different parts, the grading criteria must break down the test item into smaller items.

If using the computer to analyze the results, the following guidelines may be helpful. After the essay item is graded, the instructor records the correct response on the answer sheet.

If the grading criteria is either correct or incorrect, then the answer sheet is marked either "a" for correct or "b" for incorrect. The computer will record the number of correct and incorrect.

If a numerical grade is used, the grading criteria should indicate how partial credit is assigned. On the answer sheet the letter corresponds to a number printed above/below it. For example, "a" corresponds to 1, "b" corresponds to 2, etc. If the student can receive partial credit, the instructor marks the letter that corresponds to the amount of credit given.
• For courses that calculate the difficulty index by hand, refer to Figure 8-1 below.

<table>
<thead>
<tr>
<th># of Exams</th>
<th>Point Value</th>
<th>Exam Totals</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Sum</th>
<th>Dif</th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
<td>5</td>
<td>70</td>
<td>2</td>
<td>10</td>
<td>10</td>
<td>16</td>
<td>20</td>
<td>12</td>
<td>218</td>
<td>.62</td>
</tr>
</tbody>
</table>

**FIGURE 8-1 ITEM DIFFICULTY FOR ESSAY ITEMS**

• Record the results for all students in the "Exam Totals" row. In this example, this number will equal 70. (2 + 10 + 10 + 16 + 20 + 12 = 70). In this example, 2 students received no credit, 10 students received 1 point, 10 students received 2 points, etc.

• Multiply each column in the EXAM Totals row by the actual point achieved (not overall point value). Add the results together and record in the "Sum" Column:

\[(0 \times 2) + (1 \times 10) + (2 \times 10) + (3 \times 16) + (4 \times 20) + (5 \times 12) = 218\]

Multiply the total number of answer sheets times the actual point value of the test-item. For this example it is 70 X 5.

Determine the item difficulty as follows:

\[
\text{SUM} = \frac{(70 \times 5)}{218} = \frac{350}{218} = .62
\]

8.4. Analyzing Performance Tests

• Analyzing the results of a performance test can also provide the course managers with valuable information on how to improve training. For example, it may be determined through the analysis that students do not perform well on a particular step on the job sheet. The job sheet, the instructor or the lab should be reviewed to determine possible causes. Corrective action can be taken as described for analyzing knowledge tests. The process of
analyzing results of a performance test may be similar to that of the essay test items.

- First establish grading criteria. Next record the results on an answer sheet if the results are to be computer analyzed. If the performance test is divided into steps, each step can be analyzed. As with essay items, the grade may be a numerical grade or SAT/UNSAT. In either case, the results can be analyzed.

8.5. Test Analysis

- Analyzing individual test items is only part of the analysis process. Statistically, individual test items may appear acceptable, but when placed together on the test, may not accomplish what the test was designed to do. This section deals with guidelines for conducting test analysis.
- Content validity is the extent to which a test measures the objectives. Tests should have content validity prior to the conduct of test analysis. To ensure a test has content validity, review the test design and verify it for accuracy. For example, are the items written to a knowledge level appropriate with the objective? Are there enough items to adequately measure the objective?
- Measures of central tendency are statistical measures commonly referred to as the mean, median and mode and are used to help analyze test results. For the purpose of this handbook, the only one discussed is the mean.
- The mean is the average score. To calculate the mean add all the scores and divide the sum by the number of scores. Since the mean uses all the scores, it is sensitive to the extremely high and low scores. To determine the mean, use at a minimum 100 answer sheets. Even with a sample this size, the mean may need to be recalculated if additional classes vary a great deal.
- Studying the mean can provide two important pieces of information. One has to do with the level of difficulty between versions of a test. Recall that all versions of a test are to be of equal degrees of difficulty. If the mean has been established at 80 and version 1 has a mean of 81 and version 2 has a mean of 88, the two tests do not appear to have the same degree of difficulty.
• The second has to do with student performance. If the students' scores vary a great deal around the mean, this may be some indication of a problem. For example, are students being assigned to class who do not meet course prerequisites? If the mean is high, are there students who need to be accelerated? If the mean is very low, can remedial instruction be provided to raise the grades? Are the graduation averages consistent with expected results? Studying these statistics does not provide the answers to these problems rather they identify possible problem areas that may be investigated.

• Frequency of testing. When analyzing the tests it may be necessary to look at the frequency at which tests are administered. Testing too frequently increases the student’s opportunity for failure and encourages rote memorization, while not testing often enough prevents the early detection of problems. The recommended testing interval is a progress test for every 40-50 periods of instruction. This time interval may vary based on the complexity of the material. For example, if the material is building block material that must be learned before moving on to another concept, it may be appropriate to test this as stand alone subject matter and not wait for a weeks worth or material.

• Test item analysis results can also be used collectively to look at the test. The reviewer should look not just at the difficulty of a single item but also at the difficulty of all of them. An item may fall within the acceptable range but when all the items on a test are reviewed, the test may be too difficult or too easy. For example, several items may have acceptable difficulty ranges of .55 to .60. If the majority of the test items fall into this range, the test as a whole may be too difficult. The mean can provide some of this information but the extreme scores affect it. Another technique would be to look at the difficulty level by objectives. If the tests have been designed per the guidelines provided in earlier chapters, an analysis by objectives is possible. With this type of analysis, the reviewer can determine the objectives with which students are having difficulty.
8.6. Training Quality Indicators (TQI): TQIs, as defined in the NAVEDTRA 135 (series), are functions that, when monitored, provide the command with valuable information concerning the quality of training. The LSO, training departments and NITRAS personnel are jointly responsible for compiling this data and summarizing it for the commanding officer. This information is particularly useful when studying trends. The functions most helpful for studying the effectiveness of the testing program are Academic Review Boards, student critique program, remediation, retesting, test item analysis and student pipeline management.

- Academic Review Boards. ARBs are conducted when a student is having an academic problem. Knowing how many are conducted, what areas students have had difficulty with, the recommendations made and the actual course of action can assist in making decisions about where to concentrate manpower and other resources to improve the program.
  - For example, if the summarized results of last quarters ARBs indicate that students are having difficulty in math, it may be necessary to develop remedial materials for that area.

- Student Critique Program. The student comment sheets contain several questions concerning the tests. A review of this data may indicate how students felt about the test and why they may have failed.

- Remediation and Retesting. This program may also be studied to determine what areas the students are having problems with. For example, if 70% of the students studying voltage regulators are attending remediation for a total average time of 4 hours, it may be necessary to adjust the course length by adding more time in this area. It may further be noted that of this 70%, more than half of them fail the retest. Once again, action can be taken to determine why this material is so difficult and what can be done to make it easier for the students to understand.

- Test and Test Item Analysis. This data is only helpful if all training sites/courses are using it. The TQI summary report can identify training sites/courses which are not using test and test item analysis to improve training.
• Student Pipeline Management Data. This information may be helpful in the several ways. For example, how many students have attrited academically from training this quarter? What areas caused them the greatest difficulty? How many students are being setback, from what areas and why? Both attrition and setbacks are costly actions. While it is expected that some students will be attrited and setback, it is necessary to ensure that a poor testing program is not the reason.

SUMMARY

The process of evaluating test results to determine program effectiveness is an integral part of a quality training program. To ensure that the testing program and the tests are measuring achievement of the objectives, evaluation is a must. All of the techniques discussed in this chapter may be used to accomplish this evaluation. Each situation is different and the techniques selected should reflect the needs of the course.

The following issues were discussed in this chapter: (1) Purpose of Analysis; (2) Techniques of Test Item Analysis; (3) Procedures for Analyzing Test Items; (4) Analyzing Essay Test Items; (5) Analyzing Performance Tests; (6) Test Analysis and (7) Training Quality Indicator. The last chapter will discuss the testing plan.
CHAPTER 9

TESTING PLANS
INTRODUCTION

Chapters 1 through 8 discussed the elements necessary to develop an effective testing program. Many of these elements are explained and summarized in the course testing plan. Submission guidelines and document content are discussed in this chapter. Appendix D provides a sample-testing plan.

9.1. Guidelines for Submission

- All courses that administer performance or knowledge tests to measure objectives and/or measure individual performance will have a testing plan. If tests are not used to measure the objectives or if individual performance is not evaluated, requests for waivers, complete with justification, may be submitted to NETC, Code N63 for approval. For example, some team training courses are designed to allow a team to use the equipment. Individual evaluations do not occur. Other courses may be conducted in a seminar type environment and do not use tests to measure performance.

- When a testing plan is required, it is considered a management document and will be updated anytime there is a change to the testing program. For example, if the procedures for remediation change because of a change in student throughput, the testing plan must be updated to reflect that change.

- Where NETC is the CCA, testing plans prepared during a development/revision project and those modified during the life of the course will be submitted to NETC for approval. Where NETC is the functional and not the CCA, a copy of the approved testing plan will be submitted to NETC.

- For a multi-sited course, the CCMM is responsible for preparing the testing plan and submitting it to NETC for approval. Upon approval, the CCMM will forward a copy of the testing plan, with approval letter to each participating activity. If there are requirements in the testing plan that do not apply to the site, the site may recommend a site unique change to the testing plan and submit to the CCMM for approval.

- CCMM will provide a copy of approved changes to NETC. CCMMs should strive to provide enough general information
in the testing plan to ensure standardization at the participating activities without placing undue hardships on them. Any problems that may occur between the CCMM and the participating activity will be resolved by NETC.

9.2. Format and Content

- Format and content of the testing plans may vary between sites. Local LSO instructions may amplify on the minimum requirements as listed below. Testing plans developed for NETC courses will contain the following information:

  - Course Data. Provide a statement as to the type of course (A, C, G, F, T etc.,) and length of the course.
  - Minimum Passing Grade. When a numerical grading system is used, state the minimum passing grade for the course with the rationale for any grade higher than 63. When a SAT/UNSAT grading system is used, provide justification as to the purpose for this selection. (Refer to Chapter 1 for a discussion on the different grading systems.)
  - Methods used to Measure the Objectives. A list of the types of tests and other methods used to determine if students have met the objectives. (Refer to Chapter 1 for a list of test types and other measurement methods.)
  - Schedule of Tests. A list the tests administered in the course and the objectives measured by each test. This information must match that provided in the master schedule. For the objectives not covered by formal tests, a brief explanation as to how the objectives are measured is required. For example, objectives 1.4, 1.6 and 1.8 are measured through homework assignments and quizzes during unit 1. This is necessary to ensure that all objectives are being measured. (Refer to Chapter 1.)
  - Progress Test Procedures. Include procedures for administration, review, remediation, retest and ARBs as appropriate. (Refer to Chapter 7 for guidance in this area.)
  - Explanation of how Grades are Determined for Performance Tests. This requires a general explanation of the grading criteria for performance tests. A copy of the checklist/rating scale and grading criteria may be adequate for courses with few performance tests. Courses with a...
SAT/UNSAT grading criteria will also provide an explanation of how the grade is determined. (Refer to Chapter 3 for guidance in this area.)

- Testing Constraints are any situation that prevents the measuring of the objectives to the level stated. Testing constraints may include manpower, equipment, space, etc. If there are constraints, explain what action has been taken to correct the problem. Also include in this section justification for any policy or procedure that cannot be performed as required. If there are no constraints, state NONE.

- Grading and Weighting Criteria. Grades used in computing the student's final grade should consist of grades from practical work, progress tests, oral tests and/or comprehensive tests. When calculating the final grade, first determine the unit grade. Unit grades are determined by practical work, progress tests and within-course comprehensive tests, as applicable. Practical work normally does not exceed 10% of a unit grade. When the unit requires both knowledge and performance progress testing, the percentages assigned to the grade will be based on the situation. For example, if the unit consists primarily of knowledge with little performance, then a higher percentage of the students' unit grade should be from knowledge. If the unit consists primarily of performance then the reverse is true. When a within-course comprehensive tests is used, it will be counted as part of the unit grade.

- To compute the final grade, grades from all units will be combined to provide a single unit grade and that grade will be combined with the final comprehensive test grade. The unit grades will provide 50% to 90% of the student's final grade. Exceptions to the percentages established in these paragraphs must be justified in the testing plan. When the course is graded SAT/UNSAT, this section is not required in the testing plan.

- Test and Test Item Analysis. Provide a brief description of how often items and tests are analyzed and the types of indexes used. (Refer to Chapter 8 for guidance in this area).
This chapter contains guidance on how to develop a testing plan and what the document is to contain. Do not attempt to develop a testing plan by using only this chapter as a reference. Guidelines for a testing program are discussed throughout this handbook. Refer to each chapter as appropriate during testing plan development. Remember the sample-testing plan is only a sample. Testing plans should be developed based on the unique situations in a particular course.
APPENDIX A

TERMS AND DEFINITIONS

Academic Review Board: A board which evaluates a student's progress and recommends a course of action when the student has difficulty achieving the objectives of the course.

Acceleration Program: A program used to evaluate a student's prior learning and experience to determine if acceleration through all or any part of training is possible. The program will also include how the evaluation takes place, i.e., what tests must be passed and to what standard.

Alternate Test Versions: A version of a test administered when more than one class is on board at a time. Alternate versions may use some of the same test items as the original test.

Analysis/Evaluation: A learning level requiring a student to understand data and relationships and interpret the results. Evaluation requires a student to make a judgment or value decision concerning procedures or solutions based on data.

Application: A learning level requiring a student to demonstrate the ability to apply information taught in class to a different situation. Application may take place as a mental skill.

Checklist: A type of evaluation instrument used when a step in a process, or a characteristic of a product, is either done or not done, absent or present. Checklists are used when there are no varying degrees of performance.

Comprehension: A learning level requiring a student to demonstrate an understanding of the material taught.
Comprehensive Test: Given at the end of the instruction (final course comprehensive test) or after large blocks of material (within course comprehensive test). Used to determine mastery of the critical objectives or retention of critical materials previously measured. May be performance or knowledge.

Constraints: Resource limitations that prevent some aspect of training from being accomplished as required. For example testing constraints are anything that prevents the measurement of the objective or content area to the level stated.

Content Areas: Group of objectives with similar content and like learning levels.

Criticality of Objectives: The process of determining which objectives are the most important or critical in a course. This information is used to design tests to ensure that the most critical objectives are performed.

Difficulty Index: Test item analysis technique that calculates the difficulty on a single test item.

Discrimination Index: Test item analysis technique that calculates the ability of a single test item to measure and identify those students who have achieved the course objectives and those who have not.

Effectiveness of Alternatives: Test item analysis technique that helps to identify alternatives that may be illogical, too easy or not plausible. (Multiple choice items only.)

Formal Retest: A written or performance test administered to students after failing the original test.

Grade: Satisfactory/unsatisfactory or numerical. Is not expressed as a percentage but is based on the grading scale.

Grading Scale: Standardized interpretation of the numerical grading system.

Inference: The process of measuring lower level objectives during the measurement of higher level objectives.
Informal Retest: Methods used to measure an objective after a test failure other than written or performance testing. This may include oral retesting.

Knowledge Test: Tests that measure students' abilities to recognize facts, recall information, comprehend principles, evaluate functions or concepts and analyze findings or results in support of a job-related skill.

Minimum Passing Grade: Minimum acceptable understanding/performance required to graduate from a course.

Numerical Grading System: Method used to express a student's grade. Grades are consistent with the established grading scale and are used for ranking purposes or when degree of accomplishment of the objectives is important.

Oral Test: Test used when job performance in the fleet requires verbal demonstration of a skill.

Performance Test: Sample work situations where students demonstrate the ability to perform a skill. This may be a hands-on skill such as troubleshooting or a mental skill such as completing reports or forms.

Practical Work: Day-to-day activities completed in the laboratory, field, classroom or homework. Examples include quizzes, problem sheets, homework assignments, practice laboratories.

Pretest: A test given prior to entry into a course or unit of instruction. It may be used in the following ways: (1) For validation purposes during the pilot course. Comparison of results between tests administered at the beginning of the course and again at the end. Useful in determining the effectiveness of the instruction. (2) To assess student's knowledge and skills for possible acceleration. (3) To assess the student's prerequisite knowledge necessary to meet entry level requirements.
Process Measurement: Type of measurement used in performance testing that evaluates each step or group of steps in a process.

Product Measurement: Type of measurement used in performance testing that evaluates a finished product.

Progress Test: A test administered during the course to determine how well the student is progressing toward the achievement of the objectives. Normally covers 40-50 periods of instruction and may be administered at any time during the course.

Quiz: Short test used by the instructor to measure achievement of material recently taught.

Rating Scale: An evaluation instrument used to measure performance levels that may vary from high to low, good to bad, etc. Rating scales by definition use numerical values assigned to each performance step or job.

Raw Score: Number of items a student must answer/perform correctly in order to pass an objective or test. The raw score must be translated to a grade.

Recall: A learning level that requires a student to remember verbatim a term, fact, rule etc.

Recognition: A learning level that requires a student to identify or select the correct term, method, rule etc., as presented in training.

Remediation: A program used to aid students in achieving objectives by providing additional instructional or practical study time. Remediation programs, both voluntary and mandatory, are normally conducted outside the normal training day.

Satisfactory/Unsatisfactory Grading System: Method used to express a student’s grade when there are no varying degrees of performance. The task is either accomplished or not accomplished.
**Test Design:** The process of developing tests that measure the objectives to the appropriate standard and learning level.

**Test Development:** The process of designing, developing and validating tests.

**Test Item Analysis:** A process used to evaluate the effectiveness of a single test item. Techniques used in technical training include the difficulty index, discrimination index and the effectiveness of alternatives method.

**Test Item Bank:** Collection of all approved test items that may be used in a course of instruction.

**Test Management:** The process of managing a testing program after it has been validated.

**Test Review:** The process of reviewing a test with the class after it is administered. Each item that is missed will be reviewed to correct student misunderstanding.

**Testing Plan:** Management Control Document that outlines the course testing program.

**Training Quality Indicators (TQI):** Functions that, when monitored, provide the command with information on the overall quality of the training provided at an activity. Training departments and LSO monitor the TQIs for possible trends. TQIs are summarized by the LSO and forwarded to the commanding officer, at a minimum, quarterly or more often if required.

**Validation:** Process of ensuring that test items are technically correct and that tests measure the objectives to the level required.
APPENDIX B

PERFORMANCE TEST EVALUATION INSTRUMENT
Performance Test Grading Checklist

Student Name: __________________________ Instructor: ___________________

Course: ___________________________________ PT #: _______________________

1. Safety: PASS/FAIL

   a. Safety is not considered in the trainee’s total point count. It is a PASS/FAIL item. If there are no safety violations the trainee is allowed to perform the fault isolation procedures to completion. Any safety violation will result in the termination of the PT and a failure recorded due to a safety violation. A grade of 69 will be awarded.

   b. Safety is comprised of three elements: Personnel, Equipment, and ESD. A trainee that potentially inflicts damage to equipment or personnel is in violation of safety. Instructors must closely observe trainees and intercede prior to actual performance of violation to prevent actual damage. Grading for ESD safety compliance is twofold: total lack of compliance, which would directly damage equipment, is graded as a safety violation; all other ESD procedural performance is graded under Procedures and Techniques’.

2. Recognition of malfunction indications: (+10 possible) ____________

   a. Any item missed or an incorrect conclusion that could cause the trainee to follow the wrong path and may require instructor assistance. Trainee must acknowledge need for assistance before given. Refer to NATE 1 for point scale.

3. Interpretation of malfunction indications: (+15 possible) ____________

   Refer to NOTE 1 for point scale.

4. Troubleshooting procedures and techniques: (+25 possible) ____________

   Adherence to documented and undocumented Fault Isolation Procedures and use of GPETE/tools.

5. Evaluate trainee for the following:

   a. Knowledge of documentation.
   b. Knowledge of equipment operation.
   c. Ability to locate/utilize test points/GPETE.
   d. Ability to utilize BITE/Test panels/etc.
   e. ESD Safety procedures.
   f. Other: ____________________________

5. Problem solution: PASS/FAIL

   a. Problem solution is a PASS/FAIL item. If the trainee fails to isolate the fault within the maximum allowable time (twice the average time for that problem), the maximum grade will be 69 (failure).
**Performance Test Grading Checklist**

6. Disassembly and replacement: (+10 possible) ____________

   a. Evaluate the trainee in the performance of the procedures in SE178-AL-MMM-020, Section 10. Refer to NOTE 1 for point scale.

7. Post repair: (+10 possible) ____________

   a. Evaluate the trainee in the performance of the procedures in SE178-AL-MMM-020, Section 7. Refer to NOTE 1 for point scale.

8. Instructor assistance: (+15 possible) ____________

   a. Instructor assistance is -5 points per occurrence with a maximum of 15 points. Assistance is only given if student first agrees he requires it.

   Remarks: ____________________________________________________________
   ____________________________________________________________
   ____________________________________________________________

9. Time computations: (+15 possible) ____________

   Record the trainee’s start and stop times. Determine PT time with the following formula. Once PT time is established, find the trainee’s score using the Point Table.

   a. Start Time: ______________________
   b. Stop Time: ______________________
   c. PT Time: ______________________
   a - b = c. ______________________

   **POINT TABLE**

<table>
<thead>
<tr>
<th>% of Average Time</th>
<th>PT Average Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up to 80%</td>
<td>15 points</td>
</tr>
<tr>
<td>Up to 84%</td>
<td>14 points</td>
</tr>
<tr>
<td>Up to 88%</td>
<td>13 points</td>
</tr>
<tr>
<td>Up to 92%</td>
<td>12 points</td>
</tr>
<tr>
<td>Up to 96%</td>
<td>11 points</td>
</tr>
<tr>
<td>Up to 100%</td>
<td>10 points</td>
</tr>
<tr>
<td>Up to 110%</td>
<td>9 points</td>
</tr>
<tr>
<td>Up to 120%</td>
<td>8 points</td>
</tr>
<tr>
<td>Up to 130%</td>
<td>7 points</td>
</tr>
<tr>
<td>Up to 140%</td>
<td>6 points</td>
</tr>
<tr>
<td>Up to 150%</td>
<td>5 points</td>
</tr>
<tr>
<td>Up to 160%</td>
<td>4 points</td>
</tr>
<tr>
<td>Up to 170%</td>
<td>3 points</td>
</tr>
<tr>
<td>Up to 180%</td>
<td>2 points</td>
</tr>
<tr>
<td>Up to 190%</td>
<td>1 points</td>
</tr>
<tr>
<td>Up to 200%</td>
<td>0 points</td>
</tr>
</tbody>
</table>

When/if trainee should reach twice the average time, the PT will be stopped and scored.
Performance Test Grading Checklist

10. If this is a retake, the highest grade possible is 70. This is to prevent the trainee from making a higher grade than those students that passed the PT successfully in one attempt. This should be thoroughly explained to the trainee.

NOTE 1

Point scale range can vary due to the significance and number of steps or procedures being graded. As a rule, the following should be consistently adhered to:

a. A minor switch or indication pertinent to procedure missed, placed in wrong position, or misinterpreted, or failure to use documentation, etc... -2 points.

b. A correct indication listed as bad or a switch not pertinent to the procedure which is thrown, etc... -1 point.

c. As in A. and B. above, but for a major item: -2 to -5 points.

d. Minor procedural problem, e.g. goes to wrong book, goes to wrong equipment or panel for switch, shows general confusion about a minor step.

e. As referenced in D. above, but for a major problem: -5 to -10 points.

f. Improper use of GPETE/special tools: -5 points
NAME: _____________________  FAULT NO: _____________________

COURSE NO: __________________  COURSE NAME: __________________

FAULT AVG. TIME: _____________  PT NO: ______________________

START TIME: _____________  STOP TIME: _____________  TOTAL TIME: _______

1. Safety
   a. Personnel
   b. Equipment/ESD

2. Recognition of Fault Indications  0 to +10

3. Interpretation of Fault Indications  0 to +15

4. T/S Procedures and Techniques  0 to +25

5. Problem Solution  Pass / Fail

6. Disassembly and Replacement  0 to +10

7. Post-Repair  0 to +10

8. Instructor Assistance  0 to +15

9. Total Time  0 to +15

10. Is this a retake? - If yes, then max score is 70.
    TOTAL GRADE: _____________

Comments:
________________________________________________________
________________________________________________________
________________________________________________________

Pass  /  Fail

0  to  +10

0  to  +10

0  to  +15

0  to  +15
**MK 86 GFCS P/T SCORING CHECKLIST AND RATING SCALE**

<table>
<thead>
<tr>
<th>Trainee Name:</th>
<th>Rank:</th>
<th>Date:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructor:</td>
<td>Rate:</td>
<td>Test ID:</td>
</tr>
<tr>
<td>Time Allotted:</td>
<td>Start Time:</td>
<td>Stop Time:</td>
</tr>
</tbody>
</table>

**DIRECTIONS:** Each trainee starts test with 100 points. Use log on reverse side to record actions/decisions of trainee. For each error of omission or commission, mark or circle the value in the Error Count columns for Evaluation Category 1 and 3-6. Add the values and enter in Points column. Add entries in Points column and enter in Sub-Total column. Add values entered in Sub-Total column to determine Total Points Lost.

<table>
<thead>
<tr>
<th>Evaluation Category</th>
<th>Error Count</th>
<th>Points</th>
<th>Sub-Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Maintenance Panel Utilization: (-45 Points max.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Operated Incorrectly</td>
<td>-3</td>
<td>-3</td>
<td>-3</td>
</tr>
<tr>
<td>b. Interpreted Incorrectly</td>
<td>-6</td>
<td>-6</td>
<td>-6</td>
</tr>
<tr>
<td>2. Symptom Recognition: (List symptoms misinterpreted or overlooked)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a.</td>
<td>-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.</td>
<td>-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Technical References: (-15 Points max.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Used Incorrectly</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>b. Interpreted Incorrectly</td>
<td>-2</td>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td>4. Test Equipment: (-15 Points max.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Used Incorrectly</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>b. Interpreted Incorrectly</td>
<td>-2</td>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td>5. Fault Isolation: (-26 Points max.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Subsystem (or Subroutine) Incorrect</td>
<td>-1</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>b. Sub-unit (or Instruction) Incorrect</td>
<td>-4</td>
<td>-4</td>
<td></td>
</tr>
<tr>
<td>c. Lowest Repairable Assembly Incorrect</td>
<td>-8</td>
<td>-8</td>
<td></td>
</tr>
<tr>
<td>6. System/Equipment Restoration: (-8 Points max.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Malfunction of Equipment/System Or Incorrect</td>
<td>-4</td>
<td>-4</td>
<td></td>
</tr>
<tr>
<td>b. Explained Steps Incorrectly</td>
<td>-4</td>
<td>-4</td>
<td></td>
</tr>
<tr>
<td>7. Safety Violation: (-100 Points max.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. First Warning</td>
<td>-10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Second Warning</td>
<td>FAIL</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Total Points Lost
- Score (100 - Total Points Lost)
- Additional Time Penalty (Maximum: 15 minutes. -1 point for each minute)
- Grade
APPENDIX C

SAMPLE TEST AND TEST ITEM VALIDATION CHECKLISTS
Knowledge Test Item Worksheet

Course Title: ____________________________ CIN: __________________

Lesson Topic Title: ____________________________

Location of Supporting Curricula: ____________________________

Objective: ___________ Learning Level: ___________

Criticality: ___________ Closed Book  □  Open Book □

Version of the Test and Item Number on which the Item is Located:

Developed by: ____________________________ Date: ____________

Approved by: ____________________________ Date: ____________

Item Type:

Test Item:

Supporting Drawing, Diagram, Table, Chart, Etc.

Test Analysis Data

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Data</th>
<th>Data</th>
<th>Data</th>
<th>Data</th>
<th>Data</th>
<th>Data</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td># Tested</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrimination Hi 27%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrimination Lo 27%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date Analyzed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date Revised</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Effectiveness of Alternatives

<table>
<thead>
<tr>
<th>Choice</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hi 27%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lo 27%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Analyzed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date Revised</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sample card is only appropriate for small courses that do not utilize test item banks stored on computers.
Evaluation Checklists for Test and Test Items

This checklist is a sample checklist that may help Testing Officers and other reviewers develop effective test and test items.

1. Course Identification:

   Title: _____________________________________  CIN: ______________

   Parts/Section/Unit Module: ___________________  Lesson Number: ______

2. Test Identification:

   Test Purpose: (Check any that apply)

   _ Knowledge Progress  _ Performance Progress
   _ Knowledge Within-Course Comprehensive  _ Oral
   _ Performance Within-Course Comprehensive  _ Open Book
   _ Knowledge Final Comprehensive  _ Closed Book
   _ Performance Final Comprehensive

3. Test Overview:

   a. Total Number of Items: _____

   b. Time Allotted to complete: _____ hrs _____ min

   c. Objectives Measured: __________________________________________

   d. Number and Types of Learning Level Used:

      _ Recognition  _ Recall  _ Comprehension
      _ Application  _ Analysis/Evaluation

   e. Number and Type of Items Used:

      _ Multiple-choice  _ True/False  _ Matching
      _ Completion  _ Essay  _ Oral

   f. Does each test in the series use the same number of items to measure the objectives?  
      ☐ Yes ☐ No

   g. Are explicit directions provided to students for taking test?  
      ☐ Yes ☐ No

   h. Are test items cross-referenced?  
      ☐ Yes ☐ No

   i. Is the difficult level (mean) of the test as a whole consistent with guidelines?  
      ☐ Yes ☐ No

   j. Is the amount of material measured on the test consistent with guidelines?  
      ☐ Yes ☐ No
k. For comprehensive tests, do they measure critical objectives and are they comprehensive? □ Yes □ No

<table>
<thead>
<tr>
<th></th>
<th>ALL</th>
<th>MOST</th>
<th>SOME</th>
<th>NONE</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>l. Are objectives current, valid and realistic?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m. Are test items arranged by objectives and in sequence of material taught?</td>
<td>*□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>n. Has criticality of objectives been determined?</td>
<td>*□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>o. Do items measure knowledge required on the job?</td>
<td>*□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>p. Do test items measure knowledge of safety?</td>
<td>*□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>q. Are test items technically accurate?</td>
<td>*□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>r. Are test items free from jargon?</td>
<td>*□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>s. Are test items easy to read, unambiguous, with vocabulary familiar to the students?</td>
<td>*□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>t. Is the level of information measured appropriate for the objectives?</td>
<td>*□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>u. Is the test printed in such a manner that all items, figures and diagrams are readable/legible?</td>
<td>*□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>v. Do test items contain unimportant information (items that cause the student to memorize irrelevant data such as numbers, dimensions, etc)?</td>
<td>*□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>w. Are cues given in one item to the correct answer in the same item or another items?</td>
<td>*□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>

*Preferred entry.
Checklist for Multiple-Choice Items

<table>
<thead>
<tr>
<th></th>
<th>ALL</th>
<th>MOST</th>
<th>SOME</th>
<th>NONE</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Stem Construction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Do stems contain all information, conditions and details required to answer the question without referring to the alternatives?</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Are the stems grammatically correct?</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Are positive statement used whenever possible?</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. Are negative words emphasized?</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. Do stems omit nonessential information?</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. Are all words, phrases, etc. that pertain to all the alternatives contained in the stem?</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g. Is the correct response cued in the stem?</td>
<td></td>
<td></td>
<td></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>2. Alternative Construction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Is there only one correct item?</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Is the vocabulary used familiar to the students?</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Are all alternatives approximately the same length and complexity and expressed in the same form?</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. Are all alternatives related, meaningful and not subject to automatic elimination?</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. Does the grammar in the alternatives agree with the construction in the stem?</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. If required are alternatives arranged in some logical order? (ascending, descending, alphabetical)</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ALL</td>
<td>MOST</td>
<td>SOME</td>
<td>NONE</td>
<td>N/A</td>
</tr>
<tr>
<td>------------</td>
<td>-----</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>-----</td>
</tr>
<tr>
<td>g. Is the position of the correct answer random?</td>
<td>☑</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>h. Is all the above or none of the above used?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☑</td>
<td>☐</td>
</tr>
<tr>
<td>i. Is negative wording used in the alternatives?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☑*</td>
</tr>
</tbody>
</table>

3. Format and Punctuation:

<table>
<thead>
<tr>
<th></th>
<th>ALL</th>
<th>MOST</th>
<th>SOME</th>
<th>NONE</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Are items correctly formatted?</td>
<td>☑</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>b. Are items correctly punctuated?</td>
<td>☑</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>c. Is the EXCEPT format used appropriately and not excessively?</td>
<td>☑</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

4. Validity:

<table>
<thead>
<tr>
<th></th>
<th>ALL</th>
<th>MOST</th>
<th>SOME</th>
<th>NONE</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Are items written to the appropriate learning level?</td>
<td>☑</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>b. Are items measuring critical information?</td>
<td>☑</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

* Preferred entry
Checklist for True/False Test Items

Total Number of test items: ___________
Total Number of True/False items used: ___________

1. Is there an equal number of True/False answers?  □ Yes □ No

2. Are the answer arranged randomly? □ Yes □ No

3. Are items correctly formatted?

   ALL  MOST  SOME  NONE  N/A

   □    □    □    □    □

4. Does each item contain a direct statement or questions?

   □    □    □    □    □

5. Are items clear and concise?

   □    □    □    □    □

6. Are items consistent with common Misconceptions of the students?

   □    □    □    □    □

7. Are items phased positively?

   □    □    □    □    □

8. Are items clearly true and false?

   □    □    □    □    □

9. Are the items measuring critical Information?

   □    □    □    □    □

10. Are the learning levels consistent with The objective?

    □    □    □    □    □

11. Do any items contain specific determiners? (all, none, never, always etc.)?

    □    □    □    □    □

12. Have any items been copied verbatim from The material?

    □    □    □    □    □

*Preferred entry
Checklist for Matching Test Items

| Total Number of Test Items:  | ______________ |
| Total Number of Matching Items Used:  | _______ |

<table>
<thead>
<tr>
<th></th>
<th>ALL</th>
<th>MOST</th>
<th>SOME</th>
<th>NONE</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Do the stems contain clear direction on How the items are to be matched?</td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>2. Are the questions placed in the left-hand Column; answers in the right-hand column?</td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>3. Are there more possible answer than Questions?</td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>4. Are answers arranged in a logical order?</td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>5. Are all the answers related to each Other?</td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>6. Are the items written to the appropriate Learning levels?</td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>7. Do the items measure critical information?</td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

*Preferred entry.
Checklist for Completion or Short Answer Test Items

<table>
<thead>
<tr>
<th>Item</th>
<th>ALL</th>
<th>MOST</th>
<th>SOME</th>
<th>NONE</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Is the wording clear and concise?</td>
<td></td>
<td>*□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>2. Are the missing words or terms measuring Critical information?</td>
<td></td>
<td>*□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>3. If required, is the degree of accuracy specified?</td>
<td></td>
<td>*□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>4. Are blanks located near or at the end of Item?</td>
<td></td>
<td>*□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>5. Is the method used to indicate missing information consistent throughout the Test? (dots and blanks)</td>
<td></td>
<td>*□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>6. Are items written to the appropriate Learning levels?</td>
<td></td>
<td>*□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>7. Do items measure critical information?</td>
<td></td>
<td>*□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>8. Are items copied verbatim from the Instructional materials?</td>
<td></td>
<td></td>
<td>□</td>
<td>□</td>
<td>*□</td>
</tr>
</tbody>
</table>

*Preferred entry?
Checklist for Essay Test Items

Total Number of Test Items: ________

Total Number of Essay Items used: ________

<table>
<thead>
<tr>
<th>Question</th>
<th>ALL</th>
<th>MOST</th>
<th>SOME</th>
<th>NONE</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Do items state clearly and precisely what Response is required.</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Have the limitations of the desired Answer been specified?</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Are items relatively short and test small amounts of information?</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Does the grading criteria for each item Clearly specify how it is to be evaluated?</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Are items written to the appropriate Learning levels?</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Do items measure critical information?</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Preferred entry.
APPENDIX D

SAMPLE TESTING PLAN
SAMPLE TESTING PLAN
FOR
TITLE OF THE COURSE
CIN

PREPARED BY
NAME OF ACTIVITY
LIST OF SITES IF APPLICABLE

DATE
(REVISED)

PREPARED FOR
CHIEF OF NAVAL TECHNICAL TRAINING

D-2
PURPOSE

The purpose of this testing plan is to establish the procedures which will be used to evaluate a student's achievement of the course objectives in title of the course. (This is a canned statement and will vary only in the title of the course.)

COURSE DATA

This course provides "A" school training and is 20 weeks in length. (Since testing plans will vary based on the above factors, it is necessary to provide information to the reviewer so that quality decisions may be made.)

MINIMUM PASSING GRADE

The minimum passing grade for this course is 63. (Only provide justification if the grade is higher than 63 or SAT/UNSAT is used. Typical justification might include the need to perform the job immediately upon arrival to the fleet or the need to possess a higher-level skill and understanding for a follow-on course in the pipeline.)

METHODS USED TO MEASURE THE OBJECTIVES

Progress through this course is measured by the following:

a. Practical Work
b. Inference
c. Progress Tests (both knowledge and performance)
d. Oral Tests
e. Comprehensive Tests (both within-course and final)

Note: (The above is a complete list of possible ways to measure objectives. List only those applicable to the course.)
SCHEDULE OF TESTS

a. Progress Test 1 - A knowledge test administered at the end of lesson 1.5 (1.1, 1.2, 1.4).

b. Progress Test 2 - A performance test administered at the end of lesson 1.7 (1.5, 1.6)

(This should contain a complete list of all tests administered in the course, when they are given and the objectives each measures.)

The following objectives in Unit 1 are measured by homework assignments: (1.3 and 1.7).

The following objectives in Unit 2 are measured by lab assignments: (2.7 and 2.9).

The following objectives in Unit 3 are measured by inference during Performance Progress Test 3: (2.3, 2.5, 2.8)

(The above section is applicable only if some objectives are measured by methods other than knowledge or performance tests.)

PROGRESS TEST PROCEDURES

Performance Progress Tests are administered individually to each student. The performance tests in Unit 1 require students to complete the steps in a process and have a minimum passing grade of 63. The performance test in Unit 4 requires students to properly perform CPR and has a minimum passing grade of 85. The grade for CPR is not used to determine the student's final grade. Students will retake the CPR test until passed.

For all other performance progress tests, a review is conducted by the instructors, individually with each student. If students fail to achieve the minimum passing grade or fail a critical step or objective, remediation will be provided at the end of the working day.

After remediation, students will be retested on the complete performance test. Students are allowed only one retest. If passed, the students will be awarded the minimum passing grade for that test. If failed, the students will be recommended to an ARB for action.
(The above example is a typical situation. Remember it is only an example. The circumstances for performance testing in a course are unique and should be explained as such.)

Knowledge Progress Tests are administered to the class as a whole. The minimum passing grade is 63. After the test is graded, students are provided a grading sheet that lists all missed items. A review of all missed test items will be conducted with the students. Since the test will be reused, tests are not returned; students are not allowed to take notes during the review and all missed items are paraphrased.

Students will be assigned mandatory remediation if the test is failed or if a critical objective is failed. Students failing an objective not considered critical will be assigned additional review material prior to the following academic day.

Students failing the complete test will be retested the following day either on the failed objectives or the test as a whole. This determination is made by the course supervisor and is based on the degree of test failure. If the retest is passed, students will be awarded the minimum passing grade. If failed, students will be recommended to an ARB for action.

Students passing the test but failing a critical objective may be administered a written or oral retest depending on the degree of objective failure and overall test performance. If the retest is passed, students will be awarded the original test grade. If failed, students may be administered one additional retest prior to convening an ARB.

Students passing the test but failing an objective not considered critical may be administered a written or oral retest. If the retest is passed, students will be awarded the original test grade. If the retest is failed, students may be placed on mandatory remediation until the objective is passed. Number of retests may vary based on the situation. Students failing this type objective are not normally referred to an ARB due to the failure.

(These are only sample situations. Describe testing procedures that will be used in the particular course.)
PRACTICAL WORK PROCEDURES

Practical work consists of quizzes administered in each unit, laboratory exercises administered in units 2, 3, 4 and 5 and homework assignments administered throughout the course.

The quizzes and homework assignments are administered to identify students that may have difficulty with the progress tests. In order to identify borderline students, the minimum passing grade for the quizzes and the homework assignments has been established at a 70. Students not achieving a 70 will be assigned remediation in either a formal/structured or informal/unstructured environment. The decision as to the type of remediation will be made by the instructor based on degree of failure.

Practical work is not retested, but students may be referred to an ARB based on unsatisfactory performance of practical work.

(Practical work is an area where flexibility can be found. Use it to assist in identifying students with problems. Few restrictions apply.)

ORAL TEST PROCEDURES

Oral tests will be administered individually to each student in a board type environment. The student must respond to three of five items correctly. Items asked each student will vary but will always have the same degree of difficulty and cover the same material. Students will receive immediate feedback as to their performance. The minimum passing grade is SAT. Students not passing the oral test will be given remediation until the oral test is passed.

(Oral tests of this type are not to be confused with oral retesting. This type of examination is normally administered when students will be expected to pass oral boards or examinations when assigned to their job in the fleet.)

COMPREHENSIVE TEST PROCEDURES

Within-Course Comprehensive Tests are both knowledge and performance and are used to measure retention of the critical objectives in the course.
The knowledge tests are administered individually to the class as a whole. The minimum passing grade is 63. The within-course comprehensive tests measure critical KOs as they apply to the P0s.

Review is conducted at the class level for all missed test items. The procedures used to prevent compromise during a review of progress tests also apply for within-course comprehensive tests. Students have already been measured on the objectives prior to the within-course comprehensive tests; therefore, students will be remediated and retested for test failures only. Students passing the retest will be assigned a grade of 63. Students failing the retest will be referred to an ARB.

(It is also acceptable to continue to remediate and retest by objectives. The nature of the test, the objectives and the situation should be the deciding factors; not the preference of the course managers.)

The performance tests are administered individually to the class as a whole. The minimum passing grade is 63. Review is conducted individually with each student to discuss any missed steps or procedures. Students failing the test or a critical element of the test will be retested on the complete test. Students passing the retest will be assigned a grade of 63. Students failing the retest will be referred to an ARB.

Final Comprehensive Test is performance only and is administered individually to the class as a whole. The minimum passing grade is 63 and all critical steps must be performed without error. Time constraints prevent remediation and retesting of the final comprehensive test; therefore, students not passing the final comprehensive test will be referred to an ARB for action.

ARB Procedures. If a student's progress is not in keeping with expected performance, an ARB may be convened anytime based on the recommendations of the instructor and approval of the course supervisor.
GRADES FOR PERFORMANCE TESTS

All grades for performance tests are evaluated using rating scales. Students are assigned grades based on the grading criteria contained in the administrator's guide. All performance tests are process tests that require the students to complete the steps in a process. Students must perform all critical steps without error and achieve a grade of 63 or higher. Safety is considered a critical step. Safety violations will result in immediate failure.

(If SAT/UNSAT grades are used, state how that grade is determined, i.e., time taken to complete the task, safety, number of required correct, etc. A sample of the checklist or rating scale may also be included.)

TESTING CONSTRAINTS

All objectives are being measured. Any deviation from policy is contained where applicable in this testing plan.

(There are many different ways to address testing constraints, the above is only an example. The important concern is to list all constraints i.e., time, equipment, space, personnel, and requests for deviation from established policy. If there are no constraints or deviations, then it should be so stated.)

TEST AND TEST ITEM ANALYSIS

Student throughput is high and the following procedures apply: (1) Test item analysis will be conducted for every 100 answer sheets until the item becomes stable. (2) Once stable, the item analysis will be conducted quarterly. For knowledge tests, all three indexes will be used. Performance tests will be analyzed by steps and will use the difficulty index only. Test items will be reviewed by the instructors each time a test administered. Any item missed by 50% of the class will be reviewed for possible errors. Tests will be analyzed by the testing officer quarterly for the identification of trends.

(This paragraph should describe general procedures. If the discrimination index is not being used based on small sample size, it should be so stated in this paragraph. Approval of the testing plan constitues approval of test and test item analysis procedures.)
GRADING/WEIGHTING CRITERIA

Unit 1

Knowledge Progress Test 1  20% x SG  =  
Knowledge Progress Test 2  25% x SG  =  
Performance Progress Test 1  45% x SG  =  
Practical Work  10% x SG  =  
Unit 1 Grade  =  

Unit 2

Knowledge Progress Test 3  15% x SG  =  
Performance Progress Test 2  35% x SG  =  
Performance Progress Test 3  40% x SG  =  
Practical Work  10% x SG  =  
Unit 2 Grade  =  

Unit 3

Knowledge Progress Test 4  25% x SG  =  
Performance Progress Test 4  40% x SG  =  
Practical Work  10% x SG  =  
Within-Course Comprehensive Test 1  25% x SG  =  
Unit 3 Grade  =  

Unit 4

Knowledge Progress Test 5  40% x SG  =  
Performance Progress Test 5  50% x SG  =  
Practical Work  10% x SG  =  
Unit 4 Grade  =  

Unit 5

Knowledge Progress Test 6  30% x SG  =  
Performance Progress Test 6  30% x SG  =  
Performance Progress Test 7  30% x SG  =  
Within-Course Comprehensive Test 2  10% x SG  =  
Unit 5 Grade  =  
All unit grades are averaged and count 70% of the student's final grade. The final comprehensive test will count 30% of the final grade.

(Divide the course into units and determine a unit grade by assigning a percentage to each method used. Weight each unit grade with the final grade to determine an overall grade. The weighting criteria should reflect the objectives of the course. For example, if the critical objectives in the course are skill, then the majority of the student's grade should be based on performance. If the performance tests are graded SAT/UNSAT and a numerical grade is assigned, the grade represents only knowledge. This may not be in keeping with the objectives. This section is not required if the minimum passing grade for the course is SAT/UNSAT.)