

20 August 2004

Training
SYSTEMS APPROACH TO TRAINING: TESTING

Summary	This pamphlet provides guidance, techniques, and implementing instructions for learner measurement instruments, test design, and the development process for the U.S. Army Training and Doctrine Command (TRADOC) courses and courseware.
Applicability	This pamphlet applies to Headquarters, (TRADOC) installations, activities, and The Army School System (TASS) Training Battalions responsible for managing or performing Training Development (TD) or TD-related functions, including evaluation/quality assurance of the training products, and institutions that present the training. It also applies to non-TRADOC agencies/organizations having Memorandums of Understanding, Memorandum of Agreement, and contracts for developing training or training products for TRADOC and TASS agencies and organizations.
Suggested Improvements	The proponent for this pamphlet is the Deputy Chief of Staff for Operations and Training (DCSOPS&T). Send comments and suggested improvements on DA Form 2028 (Recommended Changes to Publications and Blank Forms) through channels to Commander, TRADOC (ATTG-CD), 5 Fenwick Road, Fort Monroe, VA 23651-1049. Suggested improvements may also be submitted using DA Form 1045 (Army Ideas for Excellence Program (AIEP) Proposal).
Availability	This publication is distributed solely through the TRADOC homepage at http://www.tradoc.army.mil/tpubs/pamsndx.htm . It is also available on the Training Development and Delivery Directorate (TDADD) homepage at http://www-dcst.monroe.army.mil/tdaa/ .

Contents

	Paragraph	Page
<u>Chapter 1</u>		
Introduction		
Purpose	1-1	4
References	1-2	5
Explanations of abbreviations and terms	1-3	5
Systems Approach to Training (SAT) overview	1-4	5

Contents (cont)

	Paragraph	Page
Regulation, pamphlet, and job aid (JA) relationships	1-5	5
Test design and development overview	1-6	7
Chapter 2		
Foundations of Army Testing		
Army testing overview	2-1	11
Types of testing	2-2	11
Mastery learning and testing	2-3	14
Performance-oriented testing	2-4	15
Chapter 3		
Fundamentals of Test Design		
Overview of test design fundamentals	3-1	16
Purpose of tests	3-2	17
Classification of tests	3-3	18
Norm/criterion-referenced test classification	3-4	18
Performance/knowledge-based test	3-5	20
Test placement in course	3-6	23
Within-course tests	3-7	23
Pretests	3-8	26
Test design	3-9	30
Chapter 4		
Criterion-Referenced Test Development		
Criterion-referenced test development overview	4-1	42
Introduction to criterion test development	4-2	42
Criterion-referenced test characteristics	4-3	43
Turning objectives into test items	4-4	48
Sequence of development	4-5	50
Chapter 5		
Test Development Management		
Test development management overview	5-1	51
Test development project steps	5-2	52
The Course Test Development Project Plan	5-3	52
Assemble the TDT	5-4	54
Determine test policy and procedures	5-5	54
Develop/revise course testing plan	5-6	56
Develop the SEP	5-7	58
Develop and validate test items	5-8	59
Write test control measures	5-9	60
Implement test plan	5-10	61
Analyze test results	5-11	62
Test development management QC	5-12	62

Contents (cont)

	Paragraph	Page
<u>Chapter 6</u>		
Development of Performance Measurements/Tests		
Performance measurements test development overview	<u>6-1</u>	64
Introduction to performance test development	<u>6-2</u>	65
Collect documentation	<u>6-3</u>	66
Select/review/revise performance objectives	<u>6-4</u>	67
Design performance test items	<u>6-5</u>	68
Determine product or process measurement	<u>6-6</u>	70
Preparing checklists for process and product measurement	<u>6-7</u>	72
Determine scoring procedures for performance measurements/tests	<u>6-8</u>	74
Writing performance test instructions	<u>6-9</u>	78
Validation of tests/test items	<u>6-10</u>	82
Update CTP and SEP	<u>6-11</u>	86
Quality control criteria for developing performance measurements/tests	<u>6-12</u>	86
<u>Chapter 7</u>		
Development of Knowledge-Based Tests		
Knowledge-based tests overview	<u>7-1</u>	86
Knowledge test development steps	<u>7-2</u>	87
Review and revise objectives	<u>7-3</u>	89
Design knowledge/cognitive skills items	<u>7-4</u>	89
Use, selection, and construction of knowledge-based test items	<u>7-5</u>	92
Validating knowledge-based test items	<u>7-6</u>	98
Validate test items	<u>7-7</u>	98
Compiling knowledge-based test items	<u>7-8</u>	103
Quality control criteria for developing knowledge-based test items	<u>7-9</u>	103
<u>Chapter 8</u>		
Test Administration and Control		
Test administration	<u>8-1</u>	103
Controlling testing material	<u>8-2</u>	105
Conducting test reviews and providing test feedback	<u>8-3</u>	112
Quality control criteria for test administration	<u>8-4</u>	116
Appendixes		
<u>A.</u> References		117
<u>B.</u> Setting Task Standards (Passing Scores) for Tests		118
<u>C.</u> Rank Ordering Learners		120
<u>D.</u> Automated Tools for Test Development		123
<u>E.</u> Review and Revise Learning Objectives		125
<u>F.</u> Interactive Courseware Test and Measurement		136
Table List		
<u>Table 3-1:</u> Knowledge-based and performance-based test item comparison, <i>page 23</i>		
<u>Table 3-2:</u> Pretest usage policy summary, <i>page 29</i>		
<u>Table 3-3:</u> Methods and activities for types of learning outcomes, <i>page 35</i>		
<u>Table 3-4:</u> Difficulty factors, <i>page 42</i>		

Contents (cont)

- [Table 6-1](#): Major differences between written and hands-on test items, *page 66*
- [Table 6-2](#): Pass-fail point value table example, *page 76*
- [Table 6-3](#): Conclusions and actions from master/nonmaster reliability trials, *page 84*
- [Table 7-1](#): Checklist for developing matching test items, *page 95*
- [Table 7-2](#): Thresholds and actions from master/nonmaster test item analysis, *page 101*
- [Table 7-3](#): Provide feedback, *page 102*
- [Table E-1](#): Examples of verbs, *page 131*
- [Table E-2](#): Examples of condition statements, *page 133*
- [Table F-1](#): CMI Administrative and Performance Tracking Functions, *page 136*

Figure List

- [Figure 1-1](#): TD policy and guidance, *page 8*
- [Figure 1-2](#): Pamphlet organization, *page 9*
- [Figure 6-1](#): Checklist rating example, *page 72*
- [Figure E-1](#): Sample learning objective, *page 125*
- [Figure E-2](#): Sample of tasks that are not unitary, *page 127*
- [Figure E-3](#): Sample of unitary tasks, *page 127*
- [Figure E-4](#): Task components, *page 128*
- [Figure E-5](#): Equation, *page 132*
- [Figure E-6](#): Sample of complete list of conditions, *page 132*

Glossary

NOTE: The use of the masculine gender in this pamphlet includes both the masculine and feminine genders.

Chapter 1 Introduction	
1.1.	Purpose. This pamphlet provides detailed guidance in support of TRADOC Regulation (Reg) 350-70 in the following areas of the student performance measurement instruments/test design and development process, for U.S. Army Training and Doctrine Command (TRADOC) courses and courseware:
	a. Foundations and fundamentals of army testing.
	b. Criterion-referenced test (CRT) development.
	c. Test development management.
	d. Development of Course Testing Plans (CTP) and Student Evaluation Plans (SEP).
	e. Development of performance-measuring instruments.

- f. Development of knowledge-based test instruments.
- g. Implementation of measurement instruments and controls.

1-2. References. The references for this pamphlet appear in [appendix A](#).

1-3. Explanation of abbreviations and terms. Abbreviations and terms appear in the [glossary](#) of this publication.

1-4. Systems Approach to Training (SAT) overview.

a. In accordance with (IAW) [AR 350-1](#), the Army's training development (TD) process is the SAT process. The SAT process is a systematic, iterative, spiral approach to make collective, individual, and self-development education/training decisions for the Army. It determines whether or not training is needed; what is trained; who needs the training; how, how well, and where the training is presented; and the training support/resources required to produce, distribute, implement, and evaluate the required education/training products.

b. The SAT Process involves five training related phases: evaluation, analysis, design, development, and implementation. Each phase and product developed has "minimum essential requirements" to meet. [TRADOC Pam 350-70-4, appendix B](#), provides a detailed discussion of the SAT Process.

c. Training development is a vital component of TRADOC's mission to prepare the Army for war. It is the responsibility of every civilian and soldier in management and training-related roles in the TRADOC headquarters, schools, field units, and supporting contractor offices. Management, at all levels, needs to have a working knowledge of the process, and ensure its efficient implementation, to save scarce resources (that is, personnel, time, process, and unnecessary product development dollars). The SAT overview, in paragraph 1-4 of [TRADOC Pam 350-70-4](#), provides the context for producing successful TD projects.

1.5. Regulation, pamphlet, and job aid (JA) relationships. This pamphlet supports and provides procedural guidance for the policy established in TRADOC Reg 350-70. The regulation directs the use of this pamphlet in the planning and conduct of test design and development. Job aids, product templates, product samples, and other supporting documents/products support this pamphlet. The pamphlet and JAs are printable as individual files, or as a single document.

TRADOC Pam 350-70-5

Supporting job aids a. Figure 1-1 depicts the relationship of this pamphlet and supporting documents/products with TRADOC Reg 350-70.

Pamphlet organization b. Figure 1-2 shows how this pamphlet is organized. Guidance provided in some chapters supports other chapters. Refer to each of these to accomplish your particular test development project. The supporting JAs listed below are also referenced throughout the pamphlet:

- (1) [JA 350-70-5.3](#), General Guidelines for Development of All Tests.
- (2) [JA 350-70-5.5](#), Guidelines for Design of IMI (Computer-Based Training (CBT)) Tests/Test Items.
- (3) [JA 350-70-5.6a](#), Guidelines for Constructing Hands On Testing (HOTS).
- (4) [JA 350-70-5.6b](#), Example of a Performance Test Measuring a Product.
- (5) [JA 350-70-5.6c](#), Example of a Performance Test Measuring a Process.
- (6) [JA 350-70-5.6d](#), Example of a Performance Test Measuring a Process and Product.
- (7) [JA 350-70-5.6e](#), Sample Performance Test: Instructions and Checklist.
- (8) [JA 350-70-5.7a](#), Guidelines for Development: All Knowledge-Based Test Items.
- (9) [JA 350-70-5.7b](#), Guidelines for Development: Multiple Choice Test Items.
- (10) [JA 350-70-5.7c](#), Guidelines for Development: Matching Test Items.
- (11) [JA 350-70-5.7d](#), Guidelines for Development: Short Answer/Completion Items.
- (12) [JA 350-70-5.7e](#), Guidelines for Development: Essay Test Items.

(13) [JA 350-70-5.7f](#), Computation of PHI (\emptyset) Coefficient For Test Item Analysis.

(14) [JA 350-70-5.8a](#), Test Administration Checklist.

(15) [JA 350-70-5.8b](#), Ground Rules for Conducting a Test Critique.

(16) [JA 350-70-5.8c](#), Sample Sensitive Test Material Sign-out and Inventory Sheet.

(17) [JA 350-70-5.8d](#), Test Control Checklist.

1-6. Test design and development overview. Effective and efficient test design and development processes (and the associated quality control (QC) of those processes) ensure that quality measuring instruments are available (1) to determine the skills, knowledge, and performance abilities of Army personnel, and (2) to evaluate the effectiveness of military instruction.

Introduction a. Student performance measurement/test design is a critical step in the design phase of the instructional development process. During test design, construct measuring instruments that measure the learner's ability to perform Learning Objectives (LOs) to the standard prescribed in the objective. During implementation, control the test instruments designed IAW their sensitive nature, and administered IAW the test plan. Compile the learners' responses, apply the GO/NO GO criterion, and collect feedback on test performance. Throughout the process, apply QC measures, to ensure development and implementation of the best final products: the Student Evaluation Plan (SEP), and the test instruments.

The Course Testing Plan b. The primary planning for testing takes place during the development of the CTP. The CTP provides all the information a learner needs about how to determine successful course completion. The plan:

(1) Details how the course proponent determines if the student demonstrated sufficient levels of competency, to pass the specified course or training.

(2) Establishes the training completion/graduation criteria/requirements.

(3) Delineates school/course counseling and retesting policy and procedures.

(4) Describes in detail each test within the course.

TRADOC Pam 350-70-5

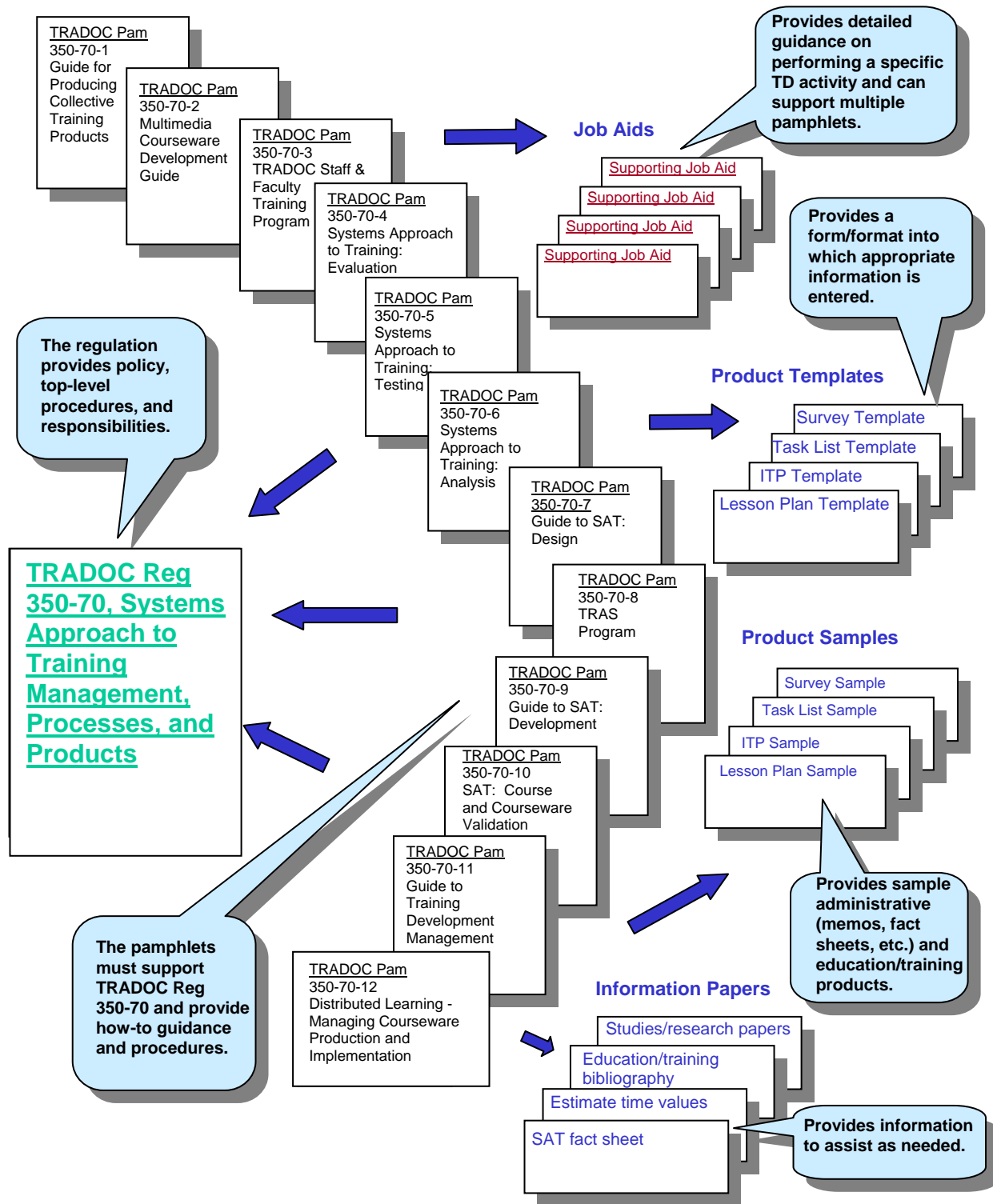


Figure 1-1. TD policy and guidance

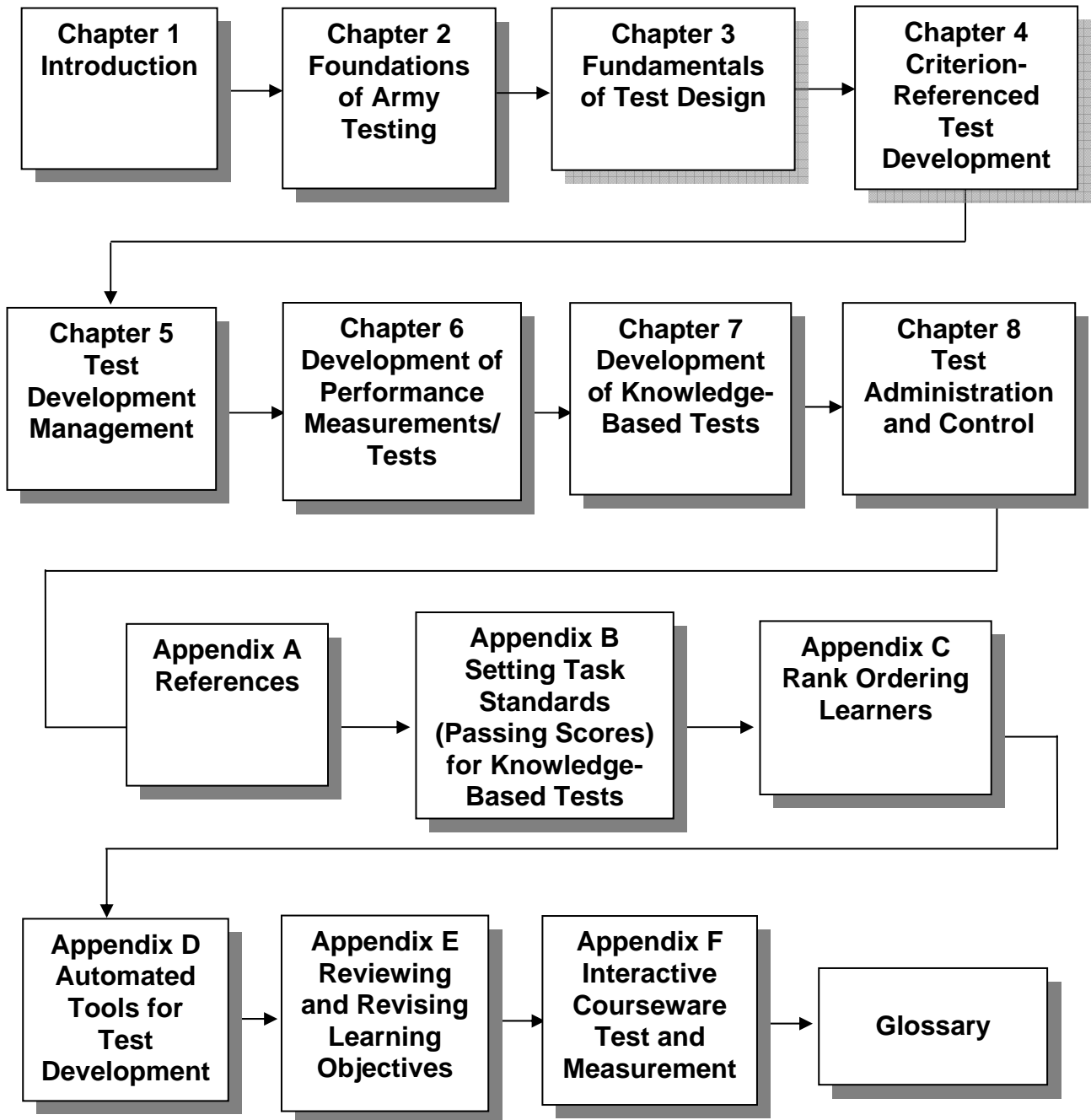


Figure 1-2. Pamphlet organization

TRADOC Pam 350-70-5

- The test plan c. After developing the general CTP, develop a test plan for each test, beginning with the performance tests. After developing the test plan for each performance test, develop the plan for each knowledge-based test.
- Test validation d. Validate the performance tests, to ensure they are administered as designed. Develop knowledge-based tests where performance tests are not feasible or necessary; usually to test supporting skills and knowledge necessary for performances the performance tests will test later in the training. Validate these knowledge-based tests, to ensure they can measure correctly and consistently the objectives they were designed to measure.
- Test implementation e. During the conduct of instruction, implement the tests IAW the SEP and the test instructions, and control the tests IAW specified guidance. Collect/observe student performance on the measuring instruments, and conduct initial analyses.
- Test evaluation f. Evaluation is a systematic, continuous process to appraise the quality (efficiency, deficiency, and effectiveness) of a program, process, or product (see [TRADOC Pam 350-70-4](#)). In the context of test instruments, collecting and analyzing data from the test administrations helps to improve the quality of the instrument. Evaluations:
- (1) Identify both intended and unintended outcomes, to enable decisionmakers to make necessary adjustments in the instructional program.
 - (2) Provide feedback used to modify the education/training program, as necessary.

Chapter 2 Foundations of Army Testing

- 2-1. Army testing overview.** This chapter relates critical foundational educational theories to Army testing. Army testing policy and procedures are built upon several key educational foundations, including criterion-referenced, performance mastery, and performance oriented tests.
- 2-2. Types of testing.** Tests are categorized into two major categories: norm-referenced tests (NRT) and CRT. These two tests differ in their intended purposes, the way in which content is selected, and the scoring process that defines how to interpret the test results.

NRT
purpose

a. The major reason for using NRT is to classify students (rank the test takers). The NRTs are designed to highlight achievement differences between and among students, to produce a dependable rank order of students, across a continuum of achievement, from high achievers to low achievers. The classic NRT distribution is a bell-shaped curve, with the scores spread out as widely as possible. To properly place students in remedial or gifted programs, instructional systems could use this classification. The NRTs, such as the Medical College Aptitude Test, are designed to reliably select the best performers.

CRT
purpose

b. In contrast to the NRT, a CRT certifies the performance of each test taker, without regard to the performance of others. Unlike the NRT, (where a test taker is defined as successful, if ahead of most of the other test takers), the CRT interpretation defines success as the ability to demonstrate specific competencies. Medical licensing board exams seek to establish criterion-referenced skills, not just rankings. (Patients want to know that the surgeon is competent, not just better than 80 percent of the graduating medical class.) There is no limit to the number of test takers succeeding on a CRT; whereas, the number of test takers selected, e.g., the top ten, the top twenty, etc., defines success on an NRT. While NRTs ascertain the rank of students, CRTs determine performance and knowledge the test takers demonstrate, not how they compare to others. The CRTs report how well students are doing, relative to a predetermined performance level, on a specified set of educational goals or outcomes, included in the total curriculum.

Army tests

c. The Army chose to use CRT, to determine how well each student learns the desired critical performances, skills, and knowledge(s), and how well the instructional system is teaching the critical tasks and supporting skills and knowledge. The purpose of classifying students is of little importance, when compared with this mandate.

TRADOC Pam 350-70-5

Definition of criterion-referenced	d. Criterion-referenced, in the “testing” context, means there is a direct and definitive link between a preestablished criterion (standard) for performance, and a test/test item that purports to measure that criterion. In criterion-referenced testing, a learner’s performance is <u>not</u> compared with another learner’s performance (this is called norm-referenced); it is only compared with the criterion.
Selection of test content - comparison of NRT and CRT	e. The choice of test content is an important difference between a NRT and a CRT. Select the content of a NRT according to how well it ranks students from high achievers to low. Determine the content of a CRT by how well it matches the learning outcomes deemed most important. Although no test can measure everything of importance, the content for the CRT is selected on the basis of its significance in the curriculum, while that of the NRT is chosen by how well it discriminates among students.
NRT interpretation	f. As mentioned earlier, a student's performance on a NRT is interpreted in relation to the performance of a large group of similar students who took the test when it was first normed. For example, if a student receives a percentile rank score on the total test of 34, this means a performance as well as, or better than, 34 percent of the students in the norm group. This type of information is useful in deciding whether or not students need remedial assistance, or are candidates for a gifted program. However, the score gives little information about what the student actually knows or can do. The validity of the score, in these decision processes, depends on whether or not the content of the NRT matches the knowledge and skills expected of the students in that particular school system.
CRT interpretation	g. It is easier to ensure the match to expected skills with a CRT. The CRTs give detailed information about how well a student performed on each of the educational goals, or outcomes, included on that test. For instance, “... a CRT score might describe which arithmetic operations a student can perform or the level of reading difficulty he or she can comprehend” (U.S. Congress, Office of Technology Assessments, 1992, p. 170). As long as the content of the test matches the content that is considered important to learn (that is, the critical tasks and supporting skills and knowledge), the CRT gives the learner, the instructor, and the Army command critical information about how much of the valued (critical) tasks content the learner can perform.
Applicability to Army testing	h. Army tests: (1) Implement criterion-referenced testing philosophy by establishing whether or not an individual can perform a task (that is, an

LO) to a preestablished standard (criterion) for performance of that task/LO. Performance is measured as a GO or NO GO against a prescribed criterion, or set of criteria - the LO standard.

(2) Are scored based upon absolute standards, such as job competency, rather than upon relative standards, such as class standings. Such concepts, used frequently in NRT, such as “averages,” “percentages,” and the “normal distribution (that is, the bell-curve)” have no applicability, relevance, or usefulness in CRT. (See [app B](#) for setting test standards.)

(3) Determine the mastery level of the learner, prior to and/or upon completion of each instructional unit (IU) of resident and distributed learning (DL) training.

(4) Standardize requirements across different target audiences (Active Component, Reserve Component, National Guard, military, civilian, etc.) to ensure uniform task mastery.

NRT versus CRT

i. It is extremely difficult to use CRTs to make norm-referenced decisions (such as comparing students with each other to determine a “commandant’s list” or an “honor graduate”). In order to make such decisions, designers frequently fall back on nonperformance-oriented tests, which are percentage-based scored. Therefore, designers should avoid this pitfall and never use the ability of a test to compare individuals to each other as a test design criterion. See [appendix C](#) for more information on how to make comparative judgments in Army courses for norm-referenced purposes.

Test theory – reducing error

j. Both CRTs and NRTs share fundamental test theory concepts. Any test score has two components: the true score and error. This is typically represented as the simple equation: $x_{\text{observed}} = x_{\text{true}} + x_{\text{error}}$. In this equation, the observed score (the test taker’s test score) consists of the two components—the true score (what is really known) and error. Error is any deviation in the score from what the test taker actually knows. Error can add to, or detract from, a true score. (Cheating is an error component that usually increases an observed score; lack of motivation is an error component that usually decreases an observed score.) The primary purpose of a systematic approach to test design—whether CRT or NRT—is to reduce the error component, so that the observed score and the true score are as nearly the same as possible. Any testing situation will always contain some error. However, minimize test error through careful attention to the systematic principles of test development and administration.

2-3. Mastery learning and testing.

a. Closely related to the criterion-referenced concept, mastery learning asserts that a learner—

(1) Tested to such a level/degree (standard), sufficient to make a definitive determination, that the learner can perform the objectives/tasks trained, within the prescribed conditions, and to the stated (mastery) standard.

(2) Tested as many times as necessary, and mastered a body of knowledge trained, within the prescribed conditions, until the mastery standard was reached. This concept is built upon the ideals of *mastery learning*.

Core idea

b. The core idea of mastery learning is that aptitude is the length of time it takes a person to learn, not how "bright" a person is, that is, everyone can learn, given the right circumstances. Adjust time to learn to fit aptitude. Also, no student is to proceed to new material until basic prerequisite material is mastered.

Mastery testing tenets

c. Regarding testing, mastery learning asserts that—

(1) Everyone may not succeed on the first try, if the material is directed at the "average" learner. (In fact, first time nonmasters are expected, not shunned. A first time "nonmaster" is not a negative event, that is, the learner is not labeled a "failure," just "not-yet" a master.)

(2) Remediation/reteaching of the material, using alternative means, methods, media, and/or material, is accomplished prior to another mastery try (retest).

(3) The test-reinstruct-retest cycle is continued until mastery is reached.

(4) The learner who "masters" the material on the second, third, or subsequent tries is NO LESS A MASTER than the one who "mastered" the material on the first try. Mastery is mastery, period.

Relationship to Army testing

d. By definition, Army tests must use the standards in LOs to distinguish between performers (masters) and nonperformers (nonmasters). Therefore, Army testing is mastery testing. Realizing that resources are limited, Army testing policy must recognize the differences in learning rates of all Army learners.

(1) Do not expect all Army learners to master the objective within a fixed length of time (that is, on the first try).

(2) Do not make an unsuccessful first try at mastery a strongly negative event for the learner.

(3) Allow, within reason, several test/reinstruct/retest cycles before learner elimination. The number of allowable cycles is variable, based upon method of instruction. Interactive Multimedia Instruction (IMI) is theoretically designed to retest an infinite number of times, until the mastery standard is obtained. However, consider resources, including alternative media/methods, human resources, material, financial resources, and time expended. See chapters [5](#) and [6](#), below, regarding development of course testing policy.

2-4. Performance-oriented testing.

- | | |
|------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Definition | <p>a. Performance-oriented testing is closely related to CRT. Performance testing relates directly (via measurement of task/skill mastery) to the performance expected of a competent individual in the actual job situation. Performance-oriented testing includes both performance tests, and tests of knowledge required to perform the tasks/skills in the actual job situation. It must ultimately determine what a person can do, not just what they might know.</p> |
| Relationship to Army testing | <p>b. All Army testing must be performance-oriented. Make the relationship between the test items, or test item set, and the performance expected on the job, clear and unambiguous. This is accomplished through—</p> <ol style="list-style-type: none"> (1) Identification of critical tasks required to perform on the job, including the expected conditions of performance, and the acceptable standard for performance. (2) Determination of the skills and knowledge required for critical task performance. (3) Development of LOs from the skills, knowledge, and critical tasks. (4) Matching the test items with the LOs (see app D, below). |

Testing to performance c. The design of Army learning is sequential and progressive. Therefore, the measure of a learners' mastery of required prerequisite skills and knowledge determines their readiness to undertake subsequent training. Determining the learner's readiness is necessary for effective, efficient, and safe instruction on the critical tasks. Sometimes, these tests are referred to as "formative tests," or, generically, as "prerequisite tests." Therefore, the relationship between some test items, measuring prerequisite or supporting skills/knowledge, and the actual task performances required on the job, is obvious, as a result of the task analysis.

Chapter 3 Fundamentals of Test Design

3-1. Overview of test design fundamentals.

a. This chapter contains an overview of test design. Subsequent chapters further provide the details necessary to design, develop, validate, and implement learner performance measures. (Quality control takes place throughout this spiral development process.) The following paragraphs include topics on:

- (1) Definitions critical to understanding the process this pamphlet describes.
- (2) Several classification methods for tests.
- (3) Guidance on use of within-course and pretests.
- (4) An overview of learning theory, as applied to the categorization and design of learner performance measures/tests.

The test design and development process

- b. The generic term "test development process" refers to:
- (1) The entire spiral development steps to design, develop, validate, implement, control, and evaluate learner performance measurement instruments/tests.
 - (2) The creation of all necessary associated documentation, such as the test development project management plan, CTP, test development plan, learner testing plan, evaluation plans and results, validation plans and results, and evaluation/data collection plans and results.

c. For organizational and reference purposes, design a test after making the decisions pertaining to number and type of tests, learning outcome expected, level of learning expected, placement in the course, number of items required, and level of optimum fidelity chosen, etc., for the course and each test/item. Document design decisions in the CTP, the SEP, and/or the test design documentation (audit trail); with certain design decisions documented in the CTP and the SEP.

d. Develop a test when constructing/writing (IAW the decisions made during the design process) validating, and approving the individual test items.

e. Implementation and evaluation follows development. Once implemented, evaluate the test (that is, collect data on test and learner performance) and use the results to determine whether or not to revise the test by reentering the process at the appropriate step. (See [chap 4](#) for additional information on the design of tests; [chaps 6](#) and [7](#) for the development (construction) of specific types of test items; and [chaps 5](#) and [8](#) for test development management, and test administration and control, respectively.)

SAT and the test development process

f. Do not confuse the design and development of a test, with the design and development phases of the SAT process. All test design, development, and validation work takes place within the context of the SAT Design Phase, although you can make needed changes anytime in the spiral SAT process.

3-2. Purpose of tests.

Primary purpose of testing

a. The primary purpose of testing is to assess learner attainment of the behaviors specified in the terminal learning objective (TLO) and enabling learning objective (ELO).

Secondary purposes of tests

- b. Tests also serve several secondary purposes, such as—
- (1) Identifying problems or weaknesses in the instruction (hopefully, during material validation). (See TRADOC Pam 350-70-10.)
 - (2) Indicating whether an individual, or class, is performing up to standards on specific objectives.
 - (3) Indicating the capability of the instructor, and the instructional medium, to facilitate learning.

- 3-3. Classification of tests.** Many schemas classify tests. The most useful of these classifications, for Army test design and development purposes, are described below. Guidance is also provided in the characteristics and use of the differing types of tests.

Types of tests

a. As introduced in [chapter 2](#), above, one major classification is a test's appropriateness for measurement and classification of learners. This classification yields the categories of NRT and CRT. (See [para 3-4](#), below.)

b. A further subtyping of CRT results in the subtypes of performance and knowledge-based (predictive) tests. This subtyping is important, because of each of the subtype's application and ability to measure different performances. These two subtypes are further described, based upon what learning outcomes they best measure (see [para 3-5](#), below).

c. An important method of subtyping is when CRTs are administered within a course. This subtyping yields pretests, within-lesson, end-of-lesson, end-of-module, end-of-phase, and end-of-course tests (see [para 3-6](#), below).

d. The last subtyping of CRT, useful in test construction, is based upon the test/item's ability to measure retention or transfer of knowledge or skills (see [para 3-7](#), below).

3-4. Norm/criterion-referenced test classification.

Test types

a. The two major types/categories of tests are—

(1) Criterion-referenced tests, which determine if learners can perform to established, well-defined training standards, or criteria (CRT are performance and knowledge-based tests).

(2) Norm-referenced tests compare a learner's performance with the performance of other learners (or the norm).

CRT

b. TRADOC and associated service schools use CRT, to determine learner competency and if the training program or lesson trains individuals to standard. A CRT—

(1) Measures an individual's ability to successfully perform the action specified in the LO. The learner's performance is compared to the LO standard.

(2) Establishes whether the learner mastered the supporting skills and knowledge required to perform the LO.

(3) Determines if the proficiency level, required for a learner to continue successfully to the next block of instruction, was met.

(4) Is scored based upon absolute standards, rather than upon relative standards, such as class standings.

(5) Provides learner scores/grades as GO (pass)/NO GO (fail).

(6) Allows classification of individual learners into two groups:

(a) Performers -- Learners who can (or are reasonably expected to) do what they were trained to do.

(b) Nonperformers -- Learners who cannot (yet) adequately do what they were trained to do.

(7) Is used as a diagnostic tool. It provides an instrument, to determine the current or entry-level performance capability of a learner. This can provide the start point for follow-on training, and allow for testing out of sections or entire courses, if the learner can demonstrate required performance mastery.

NRTs

c. Norm referenced tests measure an individual's performance against the performance of other individuals taking the same test. The NRT—

(1) Usually provides the learner's grade/score as a percentage.

(2) Does not establish if the learner can perform a specific task, or LO, to the established standard.

(3) Is useful for making relative decisions, such as which learner knows more, or who works the quickest.

(4) Is NOT used to measure learner performance in Army training.

Note: TRADOC proponent schools should test learners to determine if they can perform to established standards. They should not test learners simply to see how they compare to each other. However, refer to appendix C for detailed guidance on how to make norm-referenced decisions about learners without developing and using NRTs.

3-5. Performance/knowledge-based test.

Army training performance measurement methods
Performance test

a. For training purposes, Army CRTs are classified into two main groups: knowledge-based (sometimes called written) and performance tests. (See [chap 6](#), below, for details on the construction and use of performance tests; [chap 7](#), below, for knowledge-based tests.)

b. A performance test is one in which the learner actually performs the skill a terminal or enabling objective requires.

Clarifying written performance response format

c. What constitutes a performance test is not as clear as it may first appear. As well as testing instruments that use or simulate the use of actual equipment/situations, to perform tasks or make decisions, performance tests may seek to ascertain mastery of mental skills through written means.

(1) If a learner is required—in order to respond to a question, problem, or scenario—to mentally perform the same skill as that required on the job, the mechanism of presentation and response is not the important criteria, and the question/item is a performance item. For example, if a land navigation problem, given in written format, requires the learner to “work” through a series of steps to determine the correct answer, it is a performance item (even if the learner’s answer is captured through indicating the correct answer in writing from a choice of four alternatives). In this case, the item is performance, and the multiple-choice response is a response format/method only, and not indicative that the item is a knowledge-based (predictive) test item.

(2) The response format for a performance item is actual or simulated performance, short-answer or completion, fill-in-the-blank, or a multiple-choice format. In contrast, knowledge-based (predictive) items seek only to measure knowledge. Use the response formats of short-answer, multiple-choice, or matching for measurement purposes. (See [chap 7](#), below, for the construction of these response formats.) The learner’s ability to perform mental or physical skills or tasks (or a combination of mental and physical, known as psychomotor) is evaluated in a performance item.

Written or verbal performance tests

d. Written or verbal performance tests are conducted through writing on a piece of paper, entering into a computer, or stating orally. Use these type tests to test the following learning outcomes:

(1) Discrimination, concrete concept, and defined concept intellectual skills.

(2) Rule-learning and verbal information intellectual skills.

(3) Cognitive strategy intellectual skills.

Note: Tests that require the learner to perform a skill/task, ascertain an answer, and select from a list of possible answers are a type of performance test that has slightly less validity, due to the guessing possibility. It is best that the learner actually writes/states the answer in response, rather than just select from a list of alternatives.

Psychomotor performance tests

e. Many types of tasks, especially equipment operation tasks, involve many different intellectual and motor skills, performed in an integrated manner. Combined intellectual skills and motor skills, associated with performance of a hands-on task, are called psychomotor skills. A test that measures combined intellectual and motor skills, associated with a hands-on task, is called a psychomotor performance test. For example, the psychomotor task of bleeding a hydraulic brake system involves:

(1) Recall of a procedure (rule learning intellectual skills).

(2) The physical performance of the steps (motor skill).

(3) Recognition of the parts and tools (discrete concepts intellectual skills).

(4) Observation of the brake fluid conditions in the system (discrimination intellectual skills).

(5) Cleanliness and safety (attitude skills).

Motor skill performance tests

f. Measure motor skill performance with a written or oral test. Motor skill performance tests:

(1) Require a real or operational mock-up of equipment, or computer-generated simulations of equipment operation. (Note: If fine tactile manipulations are critical to performance, a computer-based simulation is not appropriate; use actual equipment, operational mock-up (to scale), or a simulator that accepts and responds to the necessary tactile input.)

(2) Require the learner to demonstrate mastery of an actual operational hands-on task.

(3) Have content validity. The most content-valid test of any kind of learning is an operational hands-on performance test.

TRADOC Pam 350-70-5

(4) Are generally time-consuming, because they are often conducted one-on-one, with real equipment or simulators.

Knowledge-based test uses

g. Use knowledge-based tests to predict performance in two situations:

(1) When it is not feasible to directly test the performance, test behaviors that enable performance of the desired skill. From that information, make a prediction whether the learner performs the operational task. For example, if a learner writes the steps for bleeding a brake system, there is a better probability that the learner can actually perform the task, than someone who did not know the steps. If performance testing is possible, do not use knowledge-based testing in its place.

(2) More common in a properly designed sequential and progressive training course, use knowledge-based testing to determine the learner's readiness to move forward to actual performance training and testing. That is, knowledge-based tests determine if the learner obtained certain prerequisite knowledge (defined during task analysis) necessary before actual performance is safely, efficiently, and effectively taught.

Knowledge-based tests predictions

h. Knowledge-based tests are valid to the extent that they:

(1) Predict learner performance.

(2) Measure knowledge proven necessary for task performance.

Types of knowledge-based tests

i. The most common types of knowledge-based (predictive) written test questions are essay, short answer, fill-in-the-blank, labeling, multiple-choice, matching, and true-false (although the latter is not recommended and is not addressed in this pamphlet). Computer-based knowledge-based tests use different types of input systems that have a high degree of fidelity with real-world tasks. A simple input device, such as a joystick or mouse, allows for identification by pointing with a cursor.

Comparison of performance-based test items

j. The best type of test is one that provides accurate information regarding the learner's mastery of the objective. Consider different types of test items in terms of their ability to provide the most accurate information. The differences between knowledge-based and performance test items are shown in table 3-1.

Table 3-1
Knowledge-based and performance-based test item comparison

Knowledge-based Test Item	Performance Test Item
Requires learners to demonstrate mastery of supporting knowledge, by responding to various types of written, oral, or computer-generated questions.	Requires learners to demonstrate mastery of terminal or enabling objectives , by responding to various types of written, oral, or computer-generated questions, or performing a job task under controlled conditions.
Emphasizes intellectual knowledge related to a performance objective.	Emphasizes intellectual skills associated with the hands-on performance of a motor skill (psychomotor skills).
May require learners to find, read, and use technical materials.	May require learners to find, read, and use certain technical materials (JAs, for example).
Items are intellectual skills that require mastery to enable job performance.	Items are often sequential intellectual or motor skills.
Items are independent questions, and the test item sequence does not always affect the outcome of the test.	Errors early in the performance sequence often affect the final outcome of the task.
Errors on one test item do not always affect performance on another item.	

3-6. Test placement in course. Locate course tests anywhere in a course. Normally, there is no requirement for administering a specific test at a specific point in a course. (See [para 3-9c](#), below, for more on when to test.) For discussion purposes, course tests are divided into “pretest” and “within-course” tests. See paragraphs 3-7 and [3-8](#), below, for a description of the types and uses of within-course tests and pretests.

3-7. Within-course tests.

a. The design of within course tests supports sequential, progressive training, and measures performance trained since the previous test. (They may include material from earlier training in the course, for reinforcement, etc.) They are a stand-alone lesson, or an integral part of a lesson (a learning step/activity), and may cover part of a lesson (within lesson test), one lesson (most common), or multiple lessons. Within course tests are administered end-of-course, end-of-

TRADOC Pam 350-70-5

phase, end-of-module (subcourse), end-of-lesson (most common), or within lesson.

Types, description, and usage of within-course tests

(1) End-of-course tests evaluate a learner's accomplishment of all LOs presented in the course. They are NOT required for any TRADOC-produced/managed course, and are NOT required unless there is a specific, educational requirement for that test.

(2) End-of-phase tests evaluate a learner's accomplishment of all LOs presented in the phase. They are recommended for courses structured with a significant time gap between the phases, or a major change in training focus between phases. These tests are NOT required.

(3) Use end-of-module (subcourse) tests to ensure learners can competently perform the LOs of a specific module (subcourse). They are NOT required.

(4) End-of-lesson tests are the most common type used. They measure TLO/ELOs taught within the lesson, and are required for each lesson, unless the LOs for several lessons are tested simultaneously at one administration. (**Note:** The grouping of tests measuring several TLOs is for convenience in administration only; determine TLO mastery independently for each TLO. This "group" of tests may cover several lessons, and are not necessarily used as an "end-of-module/phase" test.)

(5) Within-lesson tests are used occasionally to determine mastery of individual ELOs, or as a "graded" practical exercise. (**Note:** An ungraded practical exercise, by definition, is NOT a "test.")

End-of-DL phase test

b. A specific need for an end-of-phase test occurs when a DL phase, teaching prerequisite knowledge/skills, is closely followed by another (usually resident) phase, in which assumes, uses, and expands upon the prerequisite knowledge/skills taught in the DL phase, in the normal sequence and progression of the instruction. Use end-of-phase tests for DL phases of courses for the same purposes as any end-of-phase test.

End-of-phase test guidance

(1) Apply the following guidance when using the end-of phase tests for courses with a resident phase that follows the DL phase. While the decision to use end-of-phase tests is a design issue specific to each DL course/module, the use of an end-of-phase test is highly recommended, if the tasks/knowledge taught is not conclusively acquired. That is, the mastery level for the task/skill/knowledge taught within the DL phase is sustainable, until the time it is needed in the

resident phase (that is, their acquisition is determined via testing), within the sequential and progressive administration of the courseware (that is, if certification of competency does not take place after each lesson/module incrementally throughout the DL training), and one or more of the following conditions is true:

(a) The resident phase quickly builds upon the expected mastery of the knowledge/performances taught in the DL phase. For example, the resident phase moves quickly into the hands-on practice of procedures taught within the DL phase.

(b) There is a substantial break between the DL phase and the resident phase.

(c) The DL phase is of such length, that there is suspect of decay, or proof of decay over time, in the knowledge/performances taught early in the phase, and the need exists for reinforcement/sustainment before the learner exits the phase.

Note: This last statement might also apply to a course that is taught entirely by DL (although the test, by definition, is now an “end-of-course” test, not an “end-of-phase” test).

(d) There is no time within the resident phase for retesting or remediation.

(e) The end-of-phase test is really a “capstone” performance, or knowledge-based exercise, used to measure the mastery of the critical combination of knowledge/performances taught individually throughout the phase.

(f) There is other evidence to suggest that there is a high (rapid) decay rate for the skills/knowledge taught during the DL phase and retention determination is deemed necessary.

Determine task/skill/knowledge mastery

(2) The end-of-phase test is the last opportunity to determine task/skill/knowledge mastery and provide remediation-to-mastery, prior to the use of the knowledge within the subsequent phase. Therefore, seriously consider the use of an end-of-phase test. Some mitigation of risk is possible, by the planned pretesting of the resident phase prerequisites at the beginning of the resident phase (that is, “pretesting” those knowledge/skills acquired during the DL phase (see [para 3-8](#), below)). Nevertheless, it is more cost-effective to provide remediation and retesting during the DL phase, than to retrain/remove from training after the learner reports to the resident phase.

Note: Refer to appendix F for more detailed guidance on DL/Interactive Courseware (ICW) test and measurement.

- 3-8. Pretests.** Give a pretest before a block of instruction (lesson, phase, module, course), which serves two distinct purposes that define the types of pretests. First, use a pretest to verify if the learner previously acquired the prerequisite (entry-level) skills, knowledge, and competencies (if any) necessary, in order for the learner to master the material in the subsequent block of instruction lesson/module. This is called “prerequisite verification pretest” or “prerequisite pretest.” Secondly, use a pretest to test the learner’s prior mastery of the LOs (knowledge, skills, and competencies) the subsequent phase/module/lesson teaches(that is, for the purpose of “testing out” or reducing the objectives to master within the lesson/module/phase/course). This is called “objective mastery pretest” or “mastery pretest.” Other terms that describe this usage include “summative tests” and “mastery tests.”

Note: Sometimes the term “diagnostic test” is used interchangeably to describe either of the above types of pretests. To avoid confusion, use the appropriate name above to specify the type of pretest under discussion.

Use of prerequisite pretests

a. Prerequisite pretests.

(1) Prerequisite pretests, given at the beginning of any type of IU (that is, phase, module, or lesson) as needed, verify mastery of prerequisite objectives/tasks. If the learner’s results verify the required prerequisite skills, knowledge, and/or competencies were obtained, they proceed with the subsequent training. Action is taken if the learner does not possess necessary prerequisite skills and knowledge, which may include (in combination, where appropriate):

(a) Exclusion (not allowing learner to take course).

(b) Remediation before acceptance/entry.

(c) Conditional entry, with the simultaneous administration of remediation with new training.

(d) Conditional entry, pending proof of ability based on in-course tests.

(e) Conditional entry, based upon other evidence that the learner can reasonably master the material as expected (that is, without wasting resources on remediation).

Entry-level skill and knowledge testing

(2) Entry-level skill/knowledge testing is most important prior to the first lesson of distinct courses, phases, modules, or lessons, where the entry level skills of the different courses, phases, or modules are different; and, when there is a substantial break in time between the courses, phases, or blocks (for instance, a break of 2 months between a DL phase and a resident phase of a course, or a break of several years between functional training and advanced training in that functional area).

Use of mastery pretests

b. Mastery pretests.

(1) Mastery pretests determine the prior attainment of mastery of the tasks and/or supporting skills and knowledge (LOs) taught within the subsequent IU. It is, in fact, a version of the IU's tests/post-test, and covers the same objectives. Use objective mastery pretests before a course, phase, module, or lesson to "test-out" objectives taught during an IU. This is another way of certifying mastery.

Managing when a learner "tests-out"

(2) If the learner "tests-out" of certain instruction (especially group-paced lessons or instruction, which has a combination of self-paced and group-paced instruction), decide on the following options:

(a) Allow the learner to skip the "mastered" portion of the instruction.

(b) Move the learner to a class that is further along in the curriculum (that is, recycle forward).

(c) Give the learner advanced training.

(d) Use the learner as assistant instructor/aid/tutor.

(e) Give the learner free time.

(f) Return the learner temporarily to the unit.

(g) Give other "rewarding"-type activities.

(3) If skipping the mastered portion or recycling forward is not feasible, recommend using the learner as an assistant/aid.

Note: If the learner feels that objective mastery performance results in unrewarding/discouraging consequences, the test results may not provide a valid measure of the learner's level of mastery. As a minimum, praise learners for successful pretest objective mastery, and do not require that they take the mastered instruction.

TRADOC Pam 350-70-5

General policy for pretests

c. In accordance with TRADOC Reg 350-70, paragraph [VI-7-4e](#):

(1) Construct mastery pretests for self-paced computer-delivered training. However, recommend giving the learner an option to skip the pretest, if they desire. Base justifications for exceptions to this policy upon subject paragraph, and document.

(2) Knowledge-based, prerequisite pretests are highly recommended, in the absence of other clear and convincing evidence that the learner obtained mastery of the necessary prerequisite objectives.

(a) To avoid use of these pretests, the TD proponent assures—from knowledge based on learner records, or learner performance on previous lessons, modules, or courses—that the learner possesses the entry-level skills required.

(b) Determine and document sufficient evidence, to waive the prerequisite test requirement, on a learner-by-learner basis.

(3) If a learner is excused from taking a prerequisite test:

(a) Inform the learner that they are allowed to enter the course conditionally, based upon the evidence of attainment of prerequisites. Inform other personnel (that is, the learner's commander/supervisor) as necessary, or by local standard operating procedure (SOP) of the status of the learner.

(b) Keep a watchful eye on the learner for any failure to progress, based upon lack of prerequisites.

Performance test policy

d. It is recognized that performance pretests (either prerequisite or mastery) given to untrained personnel are sometimes dangerous, to the learner, or others. Therefore:

(1) Performance pretests are recommended (in the absence of clear evidence of prerequisite attainment) if, and only if, there exists (from the conduct of a risk assessment) a clear indication that the administration of the prerequisite tests is not harmful to personnel or equipment. In short, if harm could come to a learner or others (or equipment) when trying to perform tasks/skills in which the learner is clearly inept, do not ask the learner to perform, or stop the test immediately if testing has started.

(2) If prerequisite verification performance pretesting is not feasible, assume attainment of the performance prerequisite from less than “clear and convincing” evidence of mastery attainment. This may include knowledge-based test results, supervisor/peer/self-assessments, prior training record, etc.

(3) If mastery performance pretesting is not feasible, require the learner to go through **ALL** the training, until it is known that the learner can safely test on the task/TLO. (See [table 3-2](#) for a summary of paragraphs 3-8 and 3-9.)

Table 3-2
Pretest usage policy summary

If the pretest use is:	and the pretest is:	and:	and there is:	then pretesting is:
prerequisite verification	knowledge-based	N/A	no convincing proof of prior objective mastery	highly recommended.
	performance	performance of the objective is safely tested	no convincing proof of prior objective mastery	highly recommended.
	either knowledge-based or performance	N/A	convincing proof of prior objective(s) mastery	unnecessary.
objective mastery determination	performance	performance of the objective is safely tested	N/A	highly recommended.
	knowledge-based (assumes safe testing of performance)	subsequent instruction, which teaches the objective, is self-paced IMI	N/A	mandatory.
		subsequent instruction is not self-paced IMI	N/A	highly recommended.
for either purpose	performance	no safe testing of performance objective	N/A	not accomplished via hands-on.

Administra-
tion of
pretests

e. The two types of pretests are often administered to a learner simultaneously as a single test, or a series of tests, that measure attainment of the prerequisite objectives, as well as the prior attainment of the “to-be-taught” objectives. Each TLO, whether prerequisite or “to-be-taught,” is tested independently for mastery, based upon the test-grading criterion (cutoff/passing/mastery level). Take appropriate action, based upon results obtained for each objective:

(1) To avoid having a learner enter a later phase of course, without having the necessary phase prerequisites, test all prerequisites before entering the first phase/module/lesson (that is, do not wait to prerequisite test until just before using the prerequisite skill/knowledge, in order to allow time to plan/take mitigating action).

(2) To prevent wasting training resources, proctoring prerequisite testing before the learner reports for the planned training (that is prerequisite test at unit, DL site, or other approved location) is highly recommended (that is, test before wasting any resources). As necessary, the unit commander/delegate, or other responsible individual, should ensure the test is administered and controlled IAW chapter 8, below.

3-9. Test design.

What to measure

a. Determine what to measure.

(1) Perform an analysis of the TLO and ELO, to identify what cognitive skills and motor skills to measure.

(2) List the tasks to perform, and the TLO and ELO behaviors the test covers.

(3) Test each TLO independently of other TLOs.

(4) Adequate measurement of each TLO and ELO behavior requires one or more test items.

(5) Design tests to measure all of the cognitive and motor skills required to master each ELO and TLO behavior.

Note: This process results in determining which tests/items are performance, and which tests/items are knowledge-based.

When to test

b. Determine when to test.

(1) In general, tests are usually administered within a lesson (to determine mastery of an ELO) or after a lesson. However, you may test a logical grouping of ELOs/TLOs after a group of lessons, or at the end of a module. An end-of-phase test is usually not required, except in one instance (see [para 3-7b](#), above). The type of test (performance or knowledge-based) influences this grouping.

(2) General rules for when to test.

(a) Tests are usually given after each TLO is trained.

(b) Test TLOs simultaneously with other TLOs; however, determine learner mastery on each independent TLO tested during this “testing session.”

(c) Test TLOs sequentially if a TLO is a supporting skill/knowledge (prerequisite) for a later TLO. Test the supporting TLO (skill/knowledge) first, to ascertain the learner’s readiness for training and testing on the supported TLO.

(3) Normally, excluding retests for initial nonmastery, each TLO is tested for mastery once as a pretest, and once as a within-course test (although multiple successful repetitions of the required action during that one testing session is defined as task mastery). If you defined an accurate “mastery” standard, the learner met that standard, and the course is sequential and progressive, make the assumption that retention occurred, and allow the use of the prior obtained knowledge or skills in later portions of the course. However, you may decide to conduct another test of the same objectives, if you wish to:

(a) Reinforce the previously taught TLO(s), or

(b) Verify retention (of mastery) of the previously taught TLO.

c. Determine test length.

Test length;
coverage

(1) A test is long enough if the test (items) matches the objective, and provides sufficient information to make a master/nonmaster decision. Sometimes, one iteration of successful LO performance is sufficient to determine mastery. For other more critical TLOs, several successful iterations (or a percentage of successful versus attempts) are necessary to demonstrate true mastery. The number of TLOs tested determines knowledge-based test length. Although usually advisable if each TLO/lesson builds upon the previous TLO/lesson, each TLO does not require separate testing. A single test administration may cover and provide mastery evidence of several TLOs/ELOs.

TRADOC Pam 350-70-5

How to
decide on
the number
of items

(2) Statistically, there are a number of arguments for between 5 and 20 test items per objective. Using this advice could easily create a situation in which the test lasts longer than the course. Compromise between this idea, and more practical concerns. Generally speaking, there are five factors to help determine the number of items per objective:

(a) Consequences of misclassification. Consider the costs of judging a master as a nonmaster, or a nonmaster as a master. The greater the costs of an error, the greater the need for multiple test items.

(b) Specificity of the objective requiring testing. The more specific the objective, the smaller the number of test items needed to determine competence. This is especially true with performance tests. For example, an observer would not require a trainee to hammer a nail into a board 20 times to determine competence on this task; 3 or 4 times would suffice.

(c) Multiple TLO conditions. If the trainee is expected to perform the TLO under a number of different conditions, which might impact its performance, make decisions about which conditions to test within the learning environment.

Note: If testing under multiple conditions is not possible, multiple repetitions of performance, under the same set of conditions, brings more assurance of TLO mastery, and is recommended.

(d) Time available for testing. While an ideal test might last 1½ days for a 5-day workshop, it usually is not possible to allot such a large amount of time for testing. However, in most cases, ensure sufficient time is available, or make it available, to test each critical objective.

(e) Cost related to testing. The costs of testing should represent a balance between what it costs to pay an employee for time spent in testing, versus cost to the company due to poor job performance, resulting from inadequate identification of nonmasters. The greater the costs of poor performance, the greater the need to invest in testing.

Application and examples of test length decision-making

(3) Test length becomes a function of at least the five factors provided in paragraphs (2)(a) through (e), above. The amount of weight given to each factor varies, based upon the objective, the course, and resources. For example, a test on a very specific skill, usually performed under a single set of conditions, for which the consequences of misclassification are small, would use a single assessment for that skill. However, if assessing a very complex objective, for which the consequences of misclassification are great, and/or different conditions may affect performance, then development of multiple test items is required, based on the objective. In either situation, further decisions on test length, as a function of time and cost factors, is required. Consult subject matter experts (SMEs) in making these decisions.

Test length: time

(4) In general, unless the test coverage is many TLOs (end-of-module/phase/course), a knowledge-based test should not require more than 4-5 hours to complete (remember that it is a test of knowledge/skills, not endurance). For performance items, the test should last as long as is needed, to certify mastery, or determine nonmastery. If multiple iterations of performance are necessary to certify mastery/determine nonmastery, include a “break” between iterations. If any one iteration lasts longer than a few hours, schedule planned breaks. In only specific instances is stamina a test condition, or the learner’s stamina tested (for instance, the Army Physical Fitness Test (APFT)).

Levels of testing

d. Match desired learning levels to level of testing. In designing a test, correlate the level of testing with the level of learning found in each ELO and TLO behavior.

Bloom’s learning levels

(1) A useful taxonomy to check the match between the level of testing, and the level of learning the objective requires, is Bloom’s Taxonomy. Bloom’s Levels of Learning for the cognitive domain, from the simplest behavior, to the most complex are:

(a) Knowledge – Recall of data. Question cues: *list, define, tell, describe, identify, show, label, collect, examine, tabulate, quote, name, select, state.*

(b) Comprehension – Understand the meaning, translation, interpolation, and interpretation of instructions and problems. State a problem in one’s own words. Question cues: *summarize, describe, interpret, contrast, predict, associate, distinguish, estimate, differentiate, discuss, extend.*

(c) Application – Use a concept in a new situation, or unprompted use of an abstraction. Applies what was learned in the classroom into novel situations in the workplace. Question cues: *apply, demonstrate, calculate, complete, illustrate, show, solve, examine, modify, relate, change, classify, experiment, discover.*

(d) Analysis – Separates material or concepts into component parts, so that its organizational structure is understood. Distinguishes between facts and inferences. Question cues: *analyze, separate, order, explain, connect, classify, arrange, divide, compare, select, explain, infer.*

(e) Synthesis – Builds a structure or pattern from diverse elements. Put parts together to form a whole, with emphasis on creating a new meaning or structure. Question cues: *combine, integrate, modify, rearrange, substitute, plan, create, design, invent, what if?, compose, formulate, prepare, generalize, rewrite.*

(f) Evaluation – Make judgments about the value of ideas or materials. Question cues: *assess, decide, rank, grade, test, measure, recommend, convince, select, judge, explain, discriminate, support, conclude, compare, summarize.*

Test design
for types of
learning

(2) The outcomes of planned instruction consist of learner performances, which demonstrate acquired capabilities. The types of learning are commonly described as intellectual skills, verbal information, cognitive strategies, motor skills, and attitudes.

(a) Assess learner performance, to determine whether the newly designed instruction met its design objectives.

(b) Conduct assessment, to learn whether each learner achieved the set of capabilities the instructional objectives defined.

(3) Table 3-3 shows best methods of testing, and examples of the appropriate activities, based upon the desired outcomes (intellectual skills, verbal information, cognitive strategies, motor skills, and attitudes of the instruction).

**Table 3-3
Methods and activities for types of learning outcomes**

Type of Learning Outcome	Best Method of Testing	Activities that Indicate Achievement of Objectives
<ul style="list-style-type: none"> Intellectual Skills Discriminations 	<ul style="list-style-type: none"> Knowledge-based tests Multiple-choice Short answer 	Detect similarities or differences.
Concrete/Defined Concepts	Constructed response (labeling, sorting, matching).	Recognize examples or nonexamples.
Rule Learning	Performance of integrated tasks or constructed response (short answer).	<ul style="list-style-type: none"> Apply rule, principle, or procedure. Solve problems. Produce a product.
Verbal Information	Constructed response (fill-in-the-blank, essay questions, oral testing).	State information verbally or in writing.
Cognitive Strategies	<ul style="list-style-type: none"> Performance Tests Learner explains process to test administrator. (Oral testing) 	<ul style="list-style-type: none"> Self-report or audit trail of work done. State strategies and tactics, and expected results of actions.
Motor Skills	Performance tests	Perform smooth, timely coordinated action.
Attitudes	<ul style="list-style-type: none"> Performance tests. Observe learner in different situations. 	Display desired situated behavior.

e. Design for retention or transfer.

Designing for retention or transfer

(1) It is possible for a learner to pass a test, and still not accomplish the education or training requirement, if either the instructional program, or the test, is inadequate. The test is valid, in that it measures how well the learner *retained* the specific course material, but not how well the material is transferred. For example, a learner that remembers how to solve a particular problem in class, passes a test item requiring solution of the same problem. The test measures *retention* of course content, but the learner may not solve new problems on the job. The test did not measure how well the learner *transfers* what was learned to the job.

Retention and transfer test differences

(2) The important differences between retention and transfer tests are:

(a) Retention tests:

- Require the learner to demonstrate the retention of knowledge and skills acquired during instruction.

- Include the same examples and situations experienced in instruction.
- Require the learner to remember what was encountered during instruction.

(b) Transfer tests:

- Require the learner to demonstrate the retention of knowledge and skills acquired during instruction, and the ability to apply them to new situations and examples not encountered during instruction.
- Include different (novel) examples and situations.

Testing for retention

(3) Maintain security for the particular transfer test items the learner is given. Instructors must not teach the exact items on the test, or transfer is not inferred.

(a) For retention tests, teaching the test is not a problem.

(b) If there is only one correct way to perform the task, it is fine to “teach the test.”

(c) Give the learners the objectives at the beginning of the course.

(4) Retention tests require the learner to remember something presented in the instruction. These tests can take three forms:

Forms of retention tests

(a) Memorization. A test item requires the learner to write, state, or perform in exact terms. The learner is required to memorize exactly the content of the instruction. Any deviation is considered an error. Test item examples:

- Write the formula for water.
- State the steps for removing the fuel pump.

(b) Recall. A test item may require the learner to paraphrase, or approximate, what was taught during instruction. Test item examples:

- In your own words, define the term “discrimination.”
- Demonstrate an acceptable method for starting a car.

(c) Recognition. A test item may require the learner to look at, or read, alternatives, and recognize the correct answer. The correct answer was encountered during instruction. Test item examples:

- Which of these two fuel pumps are correctly assembled?
- Select the correct formula from this list.

Testing for transfer

(5) Transfer tests require the learner to memorize, recognize, or recall several intellectual, or motor skills, mastered during instruction, and apply these skills to new (novel) situations not encountered during instruction.

(a) For example, the learner may have to use learned rules to solve novel problems requiring the use of a formula, or using specific procedural steps.

(b) Testing for transfer is not possible, if the learner has access to the test items, and “learns” only those problems on the test.

(c) Allow the learner to practice on typical problems of this sort, prior to administration of the transfer test.

(d) The whole purpose of a transfer test is to see if the learner can apply learned intellectual, or motor skills, to novel conditions.

Sampling of complex behaviors

(6) Use transfer tests to measure complex psychomotor skills.

(a) For example, in teaching a pilot to land a plane, it is not feasible to use all possible landing strip configurations.

(b) A good transfer test would sample from the various classes of landing strip configurations, to measure a learner's ability to transfer learned psychomotor skills, to conditions not encountered in training.

Types of transfer test items and their uses

(7) The three primary types of transfer test items are:

(a) Recognition. A test item requires the learner to look at, or read, alternatives never encountered in instruction, and recognize the correct answer. Examples of recognition test items:

- Which of the following (new) examples represent negative reinforcement?
- Read the statement and select the specific answer that describes the statement.

(b) Production. A test item presents the learner with a novel practical example or situation, and asks the learner to state, or produce, the correct answer or procedure. Examples of production test items:

- Give an example of negative reinforcement not discussed in class.
- Read the case study and state the specific disorder that describes the patient.
- Select the best strategy for handling the mental patient described in the study.
- Troubleshoot an equipment malfunction not specifically covered in instruction.

(c) Application. A test item presents the learner with a novel practical problem. It asks the learner to solve the problems not encountered, using principles or procedures in instruction. Examples of application test items:

- Read this case study of a mental patient, and using principles of reinforcement, generate a resource utilization strategy for managing the patient.
- Generate tactics for landing an aircraft under conditions not encountered in instruction.
- Perceive job performance condition cues, and generate judgments as to whether a cue is an indicator of an abnormal or emergency condition, and the probable cause of the condition.

Selecting retention or transfer test items

(8) Whether you test for retention or transfer depends on the kind of behavior involved in the instructional objective.

(a) Retention tests use memorization, recall, or recognition test items. Use retention tests to measure mastery of intellectual or motor skills contained in the course of instruction.

(b) Transfer tests use recognition, production, or application test items.

Overview of transfer test design

(9) To design a transfer test for concepts mastered during instruction:

(a) Develop a list of examples and nonexamples of each concept taught in the course of instruction.

(b) The number of these examples to use in the test is based on the difficulty the learners have in learning the concept.

Testing concepts

(10) Concepts have the following characteristics:

(a) Concepts include a *class* of people, events, objects, or ideas. Members of a class share some common properties or attributes.

(b) The individual members of a class are clearly different from each other on some properties or attributes.

(c) Concepts have many examples of application. It is impossible to teach them all.

(d) To test a concept, create examples that use the concept, and then select a sample of the example to use in the test.

Attribute(s)
define
examples
and non-
examples of
a concept

(11) Base your selection of examples and nonexamples on the attributes of the members of the class of concepts, principles, etc. Some attributes are critical (that is, round objects roll). Other attributes are incidental (that is, round objects come in various colors). Examples and nonexamples for a concept are distinguished as follows:

(a) An example has the essential attributes of the concept. For example, for the concept “round,” rolling is an essential attribute. Since a ball rolls, it is an example of the concept “round.”

(b) A nonexample lacks the essential attributes of a concept, although it may share some irrelevant attribute with other members of the class. Suppose all round objects presented to teach the concept “round” were red. A red ball is an example of “round,” not because it is red, but because it rolls. A red cube would be a nonexample of round—it is red, but it does not roll.

Testing for
transfer of a
concept

(12) When testing for transfer of a concept:

(a) Ensure that students correctly make the same response to a new member of the class, which differs in some way from previously used examples of the class members. For example, if one round object shown during instructions was a phonograph record, a test item might include another example, such as a dinner plate.

(b) Ensure that students correctly make a different response to nonexamples, which share some incidental attributes with the members of a class. For example, if all the round objects presented in instruction were red, a test item might include a nonexample of a red cube.

(c) Use examples and nonexamples during instruction and in the CRT.

TRADOC Pam 350-70-5

- Advantages of using examples and non-examples
- (13) Using examples and nonexamples during instruction helps the student learn to avoid two common problems:
- (a) The student learns to include all true examples as members of the class, and is better able to transfer what was learned to the job environment.
 - (b) The student learns to exclude nonexamples from membership in the class, and is better able to transfer what was learned to the job environment.
- Selecting examples and non-examples
- (14) To prepare a list of examples and nonexamples of a concept:
- (a) Determine the critical attributes all members of the class share.
 - (b) Determine the incidental attributes that might lead students to make errors. (These are properties of the members of a class that could cause a student to incorrectly classify a nonexample as an example.)
 - (c) Prepare a list of examples and nonexamples. Use enough examples to vary each incidental attribute, and enough nonexamples to exclude each critical attribute.
 - (d) Select from the total list of examples and nonexamples those used in testing for transfer.
- Factors in transfer test development
- (15) To select a sample of examples and nonexamples from a prepared list of examples and nonexamples of a concept, first determine how large a sample is needed to test for transfer. The size of the sample depends on how difficult the concept is to learn. There are many factors that contribute to the difficulty of learning a concept; however, three are particularly relevant for developing an adequate transfer test:
- (a) The number of members of a class.
 - (b) The number of critical attributes used to describe each member of the class.
 - (c) The similarity of the critical and incidental attributes.

Determining factors in transfer test development

- (16) Consider the following determining factors to select sample size:
- (a) Determine the number of members of a class.
 - If student performance requires distinguishing among a large number of members in a class, sample more heavily than for a class having only a few members.
 - The more members there are in a class, the harder it is to see the essential similarities between them. A large class could have a dozen members.
 - (b) Determine the number of critical attributes of each member.
 - The larger the number of critical attributes the student must know, the harder it is for the student to see the essential similarities among the members of the class.
 - For example, it is harder to classify objects on the basis of size, shape, color, and texture, than on the basis of color alone. When there are more than three critical attributes, sample more heavily.
 - (c) Determine the similarity of critical and incidental attributes.
 - The more similar the critical and incidental attributes, the more difficult it is for students to identify only the correct members of the class.
 - When critical and incidental attributes are similar, sample both examples and nonexamples heavily. If critical and incidental attributes are dissimilar, sample less heavily.

Example The astronauts learned to classify minerals according to type. Suppose one objective required classifying minerals as quartz. To correctly classify sample minerals, the astronauts must understand the concept of “quartzness.” The concept involves many different kinds of quartz (members of the class). There are several critical attributes, as well, including: luster, hardness, streak, and specific gravity. The critical and incidental attributes are fairly dissimilar. (For example, the color of quartz, an incidental attribute, is not similar to any of the critical attributes).

(17) Table 3-4 depicts the difficulty factors in learning a concept, and the associated sample size.

Table 3-4
Difficulty factors

Numbers of Members in the Class	Number of Critical Attributes of Each Member	Similarity of Critical and Incidental Attributes	Number of Examples and Nonexamples to Sample
Few (<5)	Few	Dissimilar	Few
Few	Several (<5)	Dissimilar	Many
Few	Few	Similar	Moderate (5-10)
Few	Several	Similar	Many
Many (>10)	Few	Dissimilar	Few
Many	Several	Dissimilar	Moderate
Many	Few	Similar	Moderate
Many	Several	Similar	Many

Chapter 4

Criterion-Referenced Test Development

4-1. Criterion-referenced test development overview. This chapter provides general guidelines for the construction and administration of all CRT (that is, performance and knowledge-based tests). It describes critical characteristics for CRTs, and how to build in these characteristics. **Note:** Unless specified otherwise in specific sections/paragraphs, the contents of this chapter apply to both performance and knowledge-based types of CRTs.

4-2. Introduction to criterion test development.

Major requirements of test development

- a. Test development has three major requirements:
 - (1) Good tests adequately measure the instructional objectives they support.

(2) The performance required in the test should match the performance required in the objective.

(3) Prepare tests immediately after the objective is written.

Overview of the test development process

b. For test development:

(1) Review content resources to ensure that you can develop tests for all the objectives in the course of instruction.

(2) Determine the best type of test item for each objective.

(3) Develop test items for all the intellectual and motor skills in each objective.

(4) Develop retention test items for intellectual and motor skills covered in the instruction.

(5) Develop transfer test items for intellectual and motor skills not covered in instruction.

Developing the CRT

c. The process of developing the CRT involves four steps:

(1) Translating objectives into test items.

(2) Developing the CRT items.

(3) Developing objective scoring procedures.

(4) Trying out (validating) the CRT.

4-3. Criterion-referenced test characteristics.

Characteristics of tests

a. During test development, consider the following six basic characteristics, to ensure when each test is administered, it measures what is intended. The process the Army uses to ensure that a performance or knowledge-based test/item has these necessary characteristics is referred to as “validation” or the “validity and reliability” of a test/item. (See chapters [6](#) and [7](#) for specific procedures for conducting “validation” of test items.)

(1) Validity - The degree to which a test measures what it is intended to measure.

TRADOC Pam 350-70-5

(2) Reliability - The degree to which a test yields the same results consistently.

(3) Objectivity – Assurance the test is free from variations, due to factors other than the behavior measured.

(4) Comprehensiveness - The adequacy of a test to sample that which is measured.

(5) Differentiation - The ability of a test to distinguish between levels of learning.

(6) Usability - A test that is easy to administer, score, and interpret.

Definition of test validity

b. Test validity is the relevance of a test to its purpose. *Validity is the most important characteristic of a test.* For CRTs, validity refers to two characteristics of test items:

(1) The extent to which test items are direct reflections of the objectives.

(2) The adequacy with which the test items sample the objectives.

Test items as reflections of objectives

c. Preparation of criterion-referenced objectives simplifies construction of CRTs. Criterion-referenced objectives describe:

(1) The conditions of intellectual or motor skill performance.

(2) The intellectual or motor skill performance required of the learner after instruction.

(3) The accuracy and/or time standards for intellectual or motor skill performance.

Validity requirements of test items

d. A test item is valid when:

(1) It requires the learner to demonstrate the intellectual or motor skill performance stated in the objective.

(2) It requires the performance of intellectual or motor skill under the conditions stated in the objective.

(3) It is scored according to the intellectual or motor skill standards stated in the objective.

Using adequate sampling

e. Make valid not only each test item, but also the test itself.

(1) The validity of the entire test depends on how well its items sample the objectives.

(2) A test that samples one small component of a course, or unit of instruction, to the exclusion of the rest of the instructional components, is not a valid test.

Note: Sampling is not an issue for Army tests, since every TLO or ELO is tested.

Definition of test reliability

f. "Reliability" is the consistency with which a test measures the learner's mastery of the instructional objectives.

(1) If a criterion test is reliable, learners who mastered the objectives always pass, and those who did not always fail.

(2) If a criterion test is unreliable, a learner may pass, or fail, for reasons other than the ability to master the objectives.

Main factors in criterion-referenced reliability

g. The four main factors in CRT reliability are:

(1) The test itself, including general and specific test instructions, and the conditions under which the test is administered.

(2) The learner taking the test.

(3) The scoring procedures.

(4) The length of the test.

Reliability and test administration

h. For CRTs, to maximize reliability:

(1) Give the test under the most consistent conditions possible. This is the most general principle of test administration. To illustrate, suppose the national champion high school runner is chosen by learners all over the United States running once around their own outdoor high school track on 1 December. Consistent conditions would not exist for this test. Tracks are different lengths and sizes; track surfaces vary from grass to concrete; and weather conditions vary from dry to rain, sleet, or snow. The winner of the competition is not necessarily the fastest runner, but rather the one who ran under the best conditions.

(2) Make instructions to the learner as clear and simple as possible. The CRT is not a test of the learner's ability to understand complex directions.

(3) Tell the learner how the test is scored. Inform the learner whether speed or accuracy is more important, if there are penalties for errors, or if the test gives the learner simple credit for correct answers.

(4) Write all instructions, and making directions as complete as possible, without giving away answers to test items. Decide in advance how much information to give to the learner, and include this information in written instructions (see [chap 8](#) for more test administration procedures and guidelines).

Maximizing
reliability
during test
administra-
tion

i. Use the following procedures, before and during test administration, to maximize the reliability of the test results.

(1) Provide the test administrator with complete written instructions on all phases of test administration. These instructions should cover:

(a) Learner questions that are answered.

(b) Equipment and supplies needed for the test, and how they are laid out.

(c) What to do in various circumstances, such as learner illness, equipment failure, or severe weather.

(2) Provide for thorough training of the test administrator. The administrator should provide adequate supervision, to ensure tests are given, as prescribed.

(3) Make sure adequate supplies are available, and equipment is in good working order. Inconsistency in test results occurs if these factors are not addressed.

(4) Frequently inspect and calibrate the equipment and tools used for testing, to ensure consistency of operation.

(5) Protect learners from extremes of environmental conditions that might affect test scores.

Factors related to learner reliability	<p>j. The learner may be a source of unreliability. Illness, fatigue, the stress of the test, and lack of motivation may contribute to poor test scores, even if the learner mastered the objectives. Therefore, ensure the learner is rested, and treatment during the test is designed to prevent the learner from becoming <u>excessively</u> afraid of failure (some anxiety is expected, and even beneficial).</p> <p>Note: To best simulate wartime task conditions, for training purposes, on a practical exercise or performance test, generate more stressful conditions. Such generation of stress is best situationally generated via immersion in a simulation, usually, but not necessarily, of high realism (e.g., fidelity).</p>
Consistency of test scoring	<p>k. For instructor or proctor-administered tests, the scoring of tests is a major source of inconsistency. Ensure scoring is consistent from learner to learner. This is vitally important when using multiple evaluators to score learners on performance examinations. Scoring “checklists” for performance tests, or scoring keys, or automated scoring for knowledge-based (predictive) tests, significantly reduces the possibility of test inconsistency.</p> <p>Note: For learner (self) administered tests, it is essential that all instructions are absolutely clear, since the presence or availability of an instructor, or proctor, is not assured.</p>
Objectivity	<p>l. The key principle to observe in scoring is objectivity. To achieve objectivity:</p> <ol style="list-style-type: none"> (1) Set precise standards, and train the test administrator to apply them. (2) Develop scoring procedures in which subjective judgment or opinion of the scorer is not a factor. (3) Tell the test administrator exactly what to observe while scoring. (4) Clearly state the standards of performance. (5) Define successful performance so that measurements do not depend on personal judgments.
Specifying standards	<p>m. Specifying standards is essential to objectivity and reliability. Specify standards:</p>

- (1) For intellectual skills based on a single correct answer.
- (2) For psychomotor training requirements.
- (3) That indicate if a learner “did” or “did not” do a particular thing.
- (4) That indicate if a product exhibits the presence or absence of essential attributes.
- (5) That indicate if a procedure is performed within specific numerical parameters

Other ways to improve reliability

- n. Additional ways to improve reliability are to:
 - (1) Ensure measuring instruments are accurate and calibrated.
 - (2) Have several scorers score one learner, to validate scoring procedures.
 - (3) Identify the reason for any differences in scores.
 - (4) Make the standards more specific, to correct differences in scores.

Statistical indices of reliability and validity

o. Various statistical methods were devised, for obtaining numerical indices of test and test item reliability and validity. Most are not suitable for CRT, and those that are suitable, are seldom practical. For both performance and knowledge-based test items, a test item analysis technique called a master/nonmaster discrimination index proves most useful and practical, for assuring some degree of reliability of an item, and the differentiation ability of the item. Performance tests have one additional measure of reliability, consistency, and objectivity—which knowledge-based tests do not—called inter-rater reliability. (See validation sections of chaps [7](#) and [8](#), below, for procedural guidance. Other statistical reliability/validity indices are not addressed in this pamphlet.)

4-4. Turning objectives into test items.

Translating objectives into test items

- a. A valid test item is derived from an objective, that was written to describe:
 - (1) The performance required.

- (2) The conditions of performance.
- (3) The standards required for speed and/or accuracy.

Note: See [appendix E](#) for detailed guidance on the review of objectives for test item construction.

General guidelines: translating objectives into test items

b. To ensure each objective is properly translated into test items, compare each objective to the corresponding test item(s).

(1) Identify, as specifically as possible, the *inputs to the learner* (what the learner is “given”).

(2) Identify the *correct learner process and output*.

(3) Ensure the test items measure the learning behaviors and intellectual skills stated in the objectives, and the performance and measurement standards of the test items are consistent with the objective standards.

(4) Additional *inputs to the learner*, for each test item, include a description of the test item (predictive or performance) that is appropriate for measuring the objective.

(5) For a performance test item, note whether the problem involves a test of a product or process.

(6) Specify the supplies and equipment needed for the test item.

(7) Specify the *correct learner process for responding* to a test item, and the *desired outputs* for each test item, including a description of how the test item is scored.

(8) For performance tests, note what part of the learner’s performance is observed. Also note what is considered an error.

General guidelines: decisions about performance and knowledge-based test items

c. Have at least two SMEs review decisions about performance and knowledge-based test items. This ensures that the relationship between the objective and test item is as direct as possible. Ensure that:

(1) The test item requires the learner to produce the exact performance the objective requires, and no other.

(2) There are no ambiguous test item statements.

(3) The conditions under which the performance is observed are the same in the objective and the test item.

Guidelines for developing and reviewing test items

d. The major problem in developing a test item is to clearly communicate the question or problem to the learner. Develop test items, using the following general guidelines as a checklist. After developing the initial test items, a SME should review them again.

(1) Keep the language simple. The ability of the learner to comprehend difficult language ordinarily is not the skill in question.

(2) Tell the learner whether speed or accuracy is more important, and whether there are any time limits for the test, or a test item.

(3) Consider using graphics, photographs, video, audio, or other instructional media for test items, when appropriate for clear communication, or for directly relating a test item to an objective.

(4) Present the test items so they do not give the learner hints related to the correct answer.

(5) Include any instructions common to all test items in the general overall test instructions.

(6) Provide clear instructions to the test administrator. Specify what is said to the learner, and how to answer learner questions.

(7) Arrange reasonable security, to prevent learners from receiving unplanned assistance, or being disturbed while taking the test.

(8) Give clear guidance to test administrators on when to excuse a learner from a test, and under what conditions (such as equipment failure) scores are considered invalid.

Note: Checklists for the proper design of specific types of test items are found in chapters [7](#) and [8](#), below.

4-5. Sequence of development.

Performance test items first

a. Generally, in the sequence of development of test items, it is necessary to first put resources toward the development of the performance test items (from the performance objectives). Develop a

performance test on all objectives that require the development of a skill/ability, to determine a learner's level of performance on that skill/ability. Develop performance tests first, because—

(1) They are the most important to actual job task performance. They most directly test the performances (skills/abilities) closely related to actual required job performances.

(2) It provides insight into the prerequisite skills and knowledge that are trained and tested, prior to attempting training/testing on full performance. As such, their development may highlight weaknesses in the task analysis process (e.g., missing/unnecessary skills/knowledge) and missing objectives to need correction/writing.

(3) It will also assist to validate and assure the sequential and progressive sequencing of TLOs, and the effective and efficient build of skills/knowledge into full performance. Coordinate any confusion, or necessary corrections/additions, with the task analyst/course designer, before proceeding further.

Knowledge-based
second

b. After developing the performance tests (and making identified corrections to the objectives), write knowledge-based (predictive) items for those objectives that do not require performance tests (that is, for those TLOs which are pure knowledge). You may observe certain mistakes, during the development of the tests, regarding the construction or sequencing of the TLOs. Coordinate the necessary corrections with the task analyst/course designer, before proceeding further. (See [chap 7](#) on the development of knowledge-based (predictive) test items.)

Chapter 5 Test Development Management

5-1. Test development management overview. The previous chapters provide guidelines to apply when developing learner-measuring instruments. This chapter provides the procedural steps for the conduct of the test development project. It contains the organization, steps, decisions, and QC measures necessary to effectively and efficiently develop academic tests. The guidelines in previous chapters are applied during the steps indicated in this, and subsequent, chapters.

- 5-2. Test development project steps.** The primary (high-level) steps required for developing course testing documentation, procedures, and products are provided below. See [JA 350-70-5.3](#) for general guidelines on the development of all tests.

Steps in developing course academic tests and related policy

- a. Write the Course Test Development Project Plan (CTDPP) (see [para 5-3](#), below).
- b. Assemble the Test Development Team (TDT). (See [para 5-4](#), below.)
- c. Determine course testing policies and procedures. (See [para 5-5](#), below.)
- d. Write the CTP and SEP. (See paras [5-6](#) and [5-7](#), below.)
- e. Construct and validate tests/test items. (See [para 5-8](#), below.)
- f. Write test control measures. (See [para 5-9](#), below.)
- g. Implement test plan. (See [para 5-10](#), below.)
- h. Analyze test results. (See [para 5-11](#), below.)

5-3. The Course Test Development Project Plan.

Test development project plan

- a. The CTDPP is a specific type of Training Development Project Management Plan (TDPMP), designed to manage the development of learner performance measurement/testing products. For new/revised courses, you may include the CTDPP as an annex/portion of a course/course revision TDPMP.

Contents of the CTDPP

- b. As for any TD project plan, the CTDPP describes the requirements (who, what, when, where, and how) to develop tests, and the course academic testing policy and procedures necessary to implement the testing instruments.

(1) Who. List full and part-time TDT members by job/title and responsibilities, including location and contact information. If using multiple SMEs, note their respective areas of expertise. Describe the volunteers for validation trials, as specifically as possible. Satisfy necessary training requirements, before allowing an individual on the team. If not already specified, include the TD resource estimates to complete the development process. For more information, see [paragraph 5-4](#), below.

(2) What. Define the mission, scope, and expectations of the test development effort—and provide a preliminary description of the test development requirement. Initial data is provided on exactly what is expected of the team (new/revised CTP/SEP; revised/new test/items; or new administrative/test control policies/procedures). During project conduct, acquire, or update initial answers to the following questions:

- (a) Is this a new or existing course?
- (b) Why is test development/revision needed?
- (c) What data available indicates the need for test revision?
- (d) What policies and procedures already exist? Which need developing?
- (e) How many TLOs exist? Make all team members aware of the mission, scope, and expectations of the project, insofar as they are available. Questions need more complete answers, as the project developments are noted.

(3) When. Provide timelines and milestones. Since the tests are developed and validated before instructional material is validated, the timeframe required for test development, for new courses, is placed within the context of the milestone dates required for the entire course/course revision development project. For significant revisions, place the test development within the context of the timeframe when the new content is planned for implementation. Test development and validation precedes new content (material) validation.

(4) Where. Provide exact working locales of all team members, including those collocated. Maximize the use of distributed and collaborative technologies. Geographically dispersing team members, for many/most steps, is likely to result in team members with conflicting/additional responsibilities that may detract from the efficiency and effectiveness of the team.

(5) How. List the tools and other nonhuman resources available for the development project. Include references, guides, JAs, existing TRADOC and local policies and procedures, equipment, and automated hardware and software (including test authoring software). (See [app D.](#))

- 5-4. Assemble the TDT.** The CTDP identifies the team members involved in effectively and efficiently developing course testing policy and instruments. All team members must have a clear and unambiguous understanding of their roles, responsibilities, and obligations to the test development project. The primary team members include—

TDT members

- a. A minimum of one training developer (normally a GS-1750, Instructional Systems Specialist) who is specifically skilled in the process and procedures, to develop course academic tests and related policies. This individual usually functions as the project manager/team chief, also. Two or more training developers are recommended for efficiency and QC.
- b. The “content/technical” advisors, who are clear master performers (SMEs) for the tasks trained within the course. As you develop tests for varied content, include SMEs in that particular content area, on the TDT. Recommend specifying and assigning to the project at least one full-time (and at least one other part-time/on-call) SME for each content area. Verify availability for entire project, or identify and involve replacements as soon as possible (ASAP).
- c. The primary content (critical task) task analyst/TLO writer. Ensure this individual is assigned/available to the project, at least on-call/part-time.
- d. The approving authorities. Different/several levels of approval are required for different products/processes. Identify these individuals/offices in the CTDP, and notify them of their responsibilities.
- e. The master and nonmaster “volunteers” (or specific sources for the “volunteers”) for the validation trials. Identify and approve these individuals early in the test development planning process.
- f. Other team members for QC, test item review, and test validation. Bring these individuals into the process when needed. You may wish to have a representative of the local quality assurance office (QAO)/quality assurance element (QAE) as an official member, on a full or part-time basis.

5.5 Determine test policy and procedures.

Course testing policies and procedures

- a. There are critical course testing policies and procedures assessed, compiled, revised, written, and eventually provided to the learner, via the SEP. See [JA 350-70-5.5](#) for guidelines for design of IMI (CBT) tests/test items administered via DL. Some of the major areas

where policy is acquired, determined, and documented are provided below.

Note: In some areas, the proponent school and/or TDT may not have discretion (for example, if a higher headquarters regulation exists which provides policies for specific types of courses).

(1) Pretests. See [paragraph 3-8](#), above.

(2) Number of tests, retests, and remediation. The number of allowable test-remediation-retest cycles is based upon—

(a) Method of instruction (IMI) which, theoretically, is designed to retest an infinite number of times, until the mastery standard is obtained.

(b) The resources (including alternative media/methods, human resources, and time available).

(c) The resources expended to date (primarily time, as indicative of human, material, and financial investment in the learner).

(d) Time for remediation (retraining).

(e) Time between initial test and retest.

(f) Retest/recycle/retrain options and applicability.

(3) Test reclaims/feedback policy should include—

(a) Procedures for submitting (when, how, time limits).

(b) Decisionmaking processes, and documentation.

(c) Actions upon acceptance/rejection.

(4) Elimination/dismissal procedures must clearly describe procedures, appeals, and decisionmaking procedures, and documentation for each.

(5) Recognition (honor graduate/commandant's list/other) should, as a minimum, include procedures for determining honors (see [app C](#)), number of honorees, and total number of awards.

(6) Test scoring policy should include procedures, cutoffs, and mastery standard, by test.

(7) Cheating policy should include definition, learner's responsibilities, procedures, and actions.

(8) Test control (see [chap 8](#), below).

Acquiring
course
testing
policies/
procedures

b. To determine test policies and procedures:

(1) List needed policy and procedures (see [para 5-5a](#), above).

(2) Collect all existing governing policy and existing procedures mandated for the course.

(3) Obtain any waivers required.

(4) Transfer applicable policy/procedures to the CTP and SEP (or refer to the CTP and SEP, if accessible to learners).

(5) Collect/review alternative policy and procedures, from other courses/proponents, for nonmandated policy and procedures.

(6) Review/select/revise for applicability.

(7) Obtain review and approval.

(8) Transfer to the CTP and SEP.

5-6. Develop/revise course testing plan.

Content

a. In general, the CTP consists of the decisions made regarding what, how, and when the learners in the course are tested, including mastery/cutoff scores for each TLO. Unlike the SEP, it documents how and why critical decisions were made regarding learner measurement, and the source of the policies and procedures given to the learner in the SEP.

Note: It is not a "work" plan, such as the CTDPP; it is a plan which documents how the tests are implemented, and includes test policies, and test-related SOPs. Unlike the SEP, it may include sensitive testing information that requires appropriate controls (see [chap 8](#)).

- Making what, when, and how decisions
Additional “how” in CTP
- b. See chapters [3](#) and [4](#), above, for guidelines for making “what,” “how,” and “when” decisions pertaining to course tests. Summary information is provided in the paragraphs below.
- c. The most critical decisions documented in the CTP in “how-to-test” area include determining, for each TLO, whether to use performance, or knowledge-based testing; and, if knowledge-based, the exact format the test items take (essay, completion, multiple choice, matching, etc.).
- Determine performance or knowledge-based Instrument
- d. The primary determinate of whether to use performance or knowledge-based testing is the requirements (action, conditions, and especially standards) of the TLO. In making the decision, the team selects the best (most valid and reliable) method for testing the TLO, without consideration of resource constraints, or extraneous factors, such as ease of grading, or rank-ordering learners. (See chaps [3](#) and [4](#), for guidelines on making these decisions.)
- Document decisions JA – QC
- e. It is critical to document the decisions made on “what,” “when,” and “how” testing occurs. For this reason, an audit trail is extremely important, to see at a glance the critical decisions made in these areas. For QC, a completed audit trail document is recommended for each test development project, as an enclosure to the CTP.
- CTP/SEP relationship
- f. The decisions/policy first appears in the CTP, or in a direct reference from the CTP to the decision/policy (for example, an activity’s SOP). Summarize this information, rewrite as necessary, and transfer to the SEP for learner review. The SEP may refer to other documents containing test policy and procedures, as long as these documents are readily available to the learner.
- The CTP completion
- g. Sequentially, the CTP has major sections that are initially incomplete until development of the actual tests/test items, and determination of the mastery standards. “Flesh-out” these sections during the completion of the steps in the TDPMP. In some cases, sections may change as a result of test validation, material validation, initial operational trials, and subsequent implementation (including input from test critiques and analysis).
- Team involvement
- h. Involve the entire development team, in varying degrees, with the instructional developer, and the primary decisionmakers.

5-7. Develop the SEP.

SEP description a. The writing of the SEP is a direct output of the decisions made and documented in the CTP. However, the learner is only made aware of course test policy, and procedural and design decisions, not all the contents of the CTP. In appropriate sections, gear the language and content toward the learners' responsibilities.

SEP policy b. Policy regarding SEP.

(1) All courses of instruction, regardless of length, under the auspices of the proponent, requires an approved SEP.

(2) Pure informal, informational or familiarization briefings/material, that are intentionally designed without any form of learner assessment, have no need for a SEP.

(3) It is highly recommended to develop, and administer, some form of assessment, for familiarization/briefing material where the viewer (that is, learner) should retain information. At the start of these informational/familiarization events, whether or not learners are tested, inform learners of the general learning expectations, procedures, and policies.

(4) The axiom in regards to testing is: "if it is important enough to teach, it is important enough to test (that is, verify obtainment/retention)."

(5) Ensure all training designed IAW the SAT process has learner assessment instrument(s), and a corresponding SEP.

Note: In certain integrated, multiphase courses (usually for the Officer Educational System and Noncommissioned Officer Educational System), which may have common "phases/modules" and branch/military occupational specialty (MOS) specific phases/modules different proponents create, develop and give two SEPs to the learners, or integrate the SEPs into a single document. Coordination between the proponents is required, to avoid conflicting policies/procedures.

Providing SEP to learners c. Provide The SEP in writing, explain the SEP to learners during course orientation/in-processing, and apprise the learner of their responsibilities during the conduct of the course. For entirely nonresident courses, or those with initial nonresident modules/lessons, provide the Learner Assessment Plan (LAP) (and a point of contact for questions about the SEP) before course start date.

d. The SEP, as a minimum, addresses the following (usually as excerpts/summaries from the CTP):

SEP
contents

- (1) Learner performance measurement instruments and procedures administered to the learner.
- (2) The LO(s) tested at each administration.
- (3) Counseling and remedial training policy.
- (4) Reteaching/retesting policy.
- (5) Relief/recycle policy (for initial entry training courses, a flowchart depicting the relief/recycle process, to enable the learners to visualize and follow the relief/recycle procedures, is recommended).
- (6) Training graduation (pass/fail) criteria.
- (7) GO/NO GO requirements for each test.
- (8) Minimum course attendance requirements.
- (9) Definition of sustained poor performance.
- (10) Grade adjustment procedures.
- (11) Honor/distinguished graduate determination procedures (see [app C](#)).
- (12) Affiliation American Council on Education/college credit.
- (13) A listing of other nonacademic learner assessment requirements for course completion, such as APFT score, leadership assessments, peer ratings, etc.

Note: Ensure the SEP reflects TRADOC policy in each area above. Discretion is allowed only when not directly addressed by policy.

5-8. Develop and validate test items.

Assignment
by content
area

a. For this step in the CTDP, assign responsibility for individual item development to a team of instructional developers and SMEs, usually by major content area (module/group of TLOs). This team makes the remainder of critical decisions, described in chapters [3](#) and [4](#), above, in

regards to learner performance measurement methods and instruments, including test control procedures. (See [chap 6](#) for more specific guidance on performance-test development, [chap 7](#) for guidance on knowledge-based test development, and [chap 8](#) for test control and implementation procedures.)

Spiral/
simultaneous
process

b. The development and validation of test items is a spiral process, involving the steps of construction, review, validation, improvement, revalidation, etc., until the items are of sufficient quality to make instructional decisions. Also, based upon the workload of team members, simultaneously develop multiple tests covering TLOs in multiple content areas.

QC

c. In general, assure QC of the test item construction step by:

- (1) Adherence to sound test development practices IAW the guidelines in this pamphlet.
- (2) Proper test validation.
- (3) Subject matter and test design experts' test item reviews (including QAO/QAE personnel).
- (4) Decisionmaking by experts and appropriate authorities.
- (5) Documentation of decisions made, with rationale.

5-9. Write test control measures.

Respons-
ibility

a. The TD proponent is primarily responsible for:

- (1) Developing, specifying, and providing necessary and appropriate test control instructions to all instructional activities administering the test instruments. Specify exact procedures, IAW the test control measures and guidance found in [chapter 8](#), below.
- (2) Emplacing/exercising appropriate controls for sensitive materials during the development, validation, and approval process.
- (3) Ensuring the means of secure transmittal of sensitive test material to the administering activity (to include coordinating with, or providing instruction to, other activities having a distribution role).

Risk analysis b. The test instrument TD proponent should conduct a risk analysis, in order to specify the proper level of control for each test instrument/like instrument(s) during transmittal and implementation, based upon the:

(1) Type of test instrument administered (performance or knowledge-based).

(2) Instrument's method of administration (for instance, instructor/evaluator/administrator presence (live or virtually)).

(3) Target population.

(4) Prior history/experience with like populations and test instrument.

(5) Further guidance found in [chapter 8](#), below.

Note: The determination of the sensitive nature of any particular test instrument, or material, is solely at the discretion of the TD proponent.

Just enough controls c. Based upon the above, for each instrument/like instrument(s), the test proponent specifies only the minimum level of controls needed to reasonably protect only sensitive test material from unnecessary risk of compromise.

5-10. Implement test plan.

Responsibility of TD proponent a. Within this step, the TD proponent is responsible for:

(1) Providing clear test control and administration procedures to the administering activity.

(2) Preparing and providing test data collection instruments/means.

(3) Obtaining test performance data and critiques.

(4) Assessing the administration of test instruments and the control of sensitive test materials. (See [TRADOC Pam 350-70-4](#).)

Responsibility of administering activity b. The administering activity:

(1) Administers all test instruments, IAW guidance the TD proponent provides (or, if not possible, or if concerns arise, immediately contact the TD proponent for guidance).

(2) Implements the appropriate test control procedures, based upon guidance the proponent provides, and also contained in [chapter 8](#), below.

(3) Administers test reviews, IAW guidance the test proponent provides and guidance contained in this pamphlet.

(4) Administers test critique instruments, and provides data to the TD proponent.

(5) Provides recommendations/concerns for test improvement to the TD proponent.

(6) Consults the TD proponent for questions regarding reclaims.

(7) In coordination with the test proponent, develops a clear and detailed local SOP for test control, IAW [chapter 8](#), below.

5-11. Analyze test results. The TD proponent collects and analyzes test performance data, test critique data, and instructor-initiated recommendations. The TD proponent's QAO/QAE office, or other individuals/activities in the proponent's TD activity, may perform actual compilation and analysis of performance data, test critique data, and instructor-initiated recommendations. It is highly recommended, at the least, to consult the QAO/QAE for assistance in data compilation, analysis, and interpretation; and subsequent decisions about changes/improvements to test instruments, materials, and test administration processes. If the TDT no longer exists, recommend involving members of that team (primarily instructional developers and SMEs) in revisions to course testing instruments, products, and procedures.

Note: Although the actual TDT may no longer exist as a separate, defined entity, the team (in coordination with the local QAO/QAE) ensures the defining and institutionalizing of the procedures and processes, for the effective collection and analysis of data, and the implementation of subsequent changes, as described within this chapter.

5-12. Test development management QC.

TDT QC

a. Assure the competency and availability of selected team members to fulfill their role. If in doubt, double check qualifications/availability, and reassign, as necessary. Identify and approve sources of

replacements/alternates. Obtain and ensure unconditional management and team member buy-in. Develop, approve, and provide all team members the individual/role responsibilities.

Test policy and procedures

b. Exercise QC by:

(1) Timely reviews by team members, QAE/QAO personnel, and other sources of policies/procedures, as necessary.

(2) Careful documentation of the sources of policies and procedures, and rationale for additional needed policies.

(3) Obtaining necessary waivers/exceptions to policies.

(4) Decisionmaking by appropriate experts and authorities.

CTP

(5) Exercise QC of the CTP by full team involvement, review of the plan by at least two instructional developers, documentation of the source, and rationale for decisions made/recommended, and full decisionmaker approval and buy-in.

Learner assessment plan QC

c. Learner assessment plan QC procedures include:

(1) Maintaining an audit trail.

(2) Verifying that the necessary items from the CTP are addressed in the LAP.

(3) Matching the contents against LAPs from other similar courses.

(4) Quality Assurance Element/QAO personnel review.

Test control measures

d. To assure QC of the output of this step, adhere to guidance in [chapter 8](#), below; ensure review by appropriate personnel; and document the risk analysis and rationale for team decisions made, along with appropriate decisionmakers approvals.

Implementing test plan QC for TD proponent

e. The TD proponent ensures QC of this step by:

(1) Using checklists for the development of test administration procedures and guidance.

(2) Instituting appropriate level(s) of reviews and approvals during the development process (TD proponent).

(3) Assessing the implementation of test administration instructions, test controls, and data collection products and procedures by the administering activity(ies).

(4) Collecting, analyzing, and incorporating (as necessary) feedback from learners and instructors on test instruments and materials, and test administration procedures.

QC for test administering activities

f. See responsibilities listed in paragraph 5-12.e above. Also, self-assessments, by the administering activities of test control and administration procedures, will assist in QC during implementation.

Analyzing test results

g. The collection and analysis of test performance data, test critique data, and instructor-initiated recommendations is a continuous process. Competent instructional development and content experts validate recommendations resulting from this analysis, and appropriate authority approval obtained. The test development process is reinitiated/reentered where needed. For critical, immediate changes, give interim instructions, by the most expeditious means, to the administering activities, pending publication and distribution of revisions.

Chapter 6

Development of Performance Measurements/Tests

6-1. Performance measurements test development overview.

a. This chapter provides the details of the procedures to develop and validate performance (hands-on) test items (as part of constructing and validating test/test items in the test development process in chap 5, above). Make and implement specific performance test development decisions (IAW the guidelines in chapters [3](#) and [4](#)) during the development of performance test items.

Note: A “performance test” and a “performance test item” are synonymous, since each “item” is usually a separate test, measuring a single TLO.

Introduction b. It is recognized that there is a fundamental difference between knowing about a job, and the ability to perform the job. “Knowledge of” is really an essential ingredient for doing a complex job correctly, but while it is a necessary condition, it is rarely a sufficient condition for satisfactory performance. Individuals may provide a good guess on a written test, but can seldom carry off a successful deception, when a realistic performance test is required. One of the great virtues of the performance test is its impressive “face validity” and credibility—the task to perform closely resembles the job itself. Frequently, the performance is nothing more than a work sample—performing an actual task, but outside of the normal job environment. Nothing but actual on-the-job performance measures competencies as effectively as the performance test.

Mandate c. In accordance with [AR 350-1](#), all Army tests are performance-oriented (that is, either testing performance directly, or testing the knowledge necessary for performance, as determined from the task analysis). Performance tests are always preferred over knowledge-based (predictive) tests for demonstration of skill mastery.

6-2. Introduction to performance test development.

General overview of performance test development a. Performance tests are the best method for the accurate assessment of an individual’s overall competency. In order to develop good performance test items, ensure that the objectives are sufficiently specific, determine the level of test fidelity possible, determine scoring procedures, write the items and instructions, validate items and instructions, obtain approval, and update the course test plan and the SEP. Note that some of the above steps are not necessarily linear in their performance. As with almost all TD processes, use of a spiral development requires returning to previous steps, for necessary corrections and improvements.

b. See [paragraph 3-5](#), above, to review key elements of performance items.

Review of critical differences of performance and knowledge-based tests c. Performance tests require the learner to perform an overt action or series of actions, rather than verbalize or write (unless the required performance is speaking or writing). Table 6-1 shows a comparison between written and performance test items.

Table 6-1
Major differences between written and hands-on test items

Written Test Items	Performance Test Items
Primarily abstract or verbal.	Primarily nonverbal.
Items address knowledge and content.	Items are skills, performances, or job related decisions.
Items usually address independent aspects.	Items sequentially presented. Errors early in the sequence may affect later items.

Test development steps d. The test development steps are discussed in paragraphs 6-3 through 6-5, and 6-9 through 6-11, below.

6-3. Collect documentation.

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

b. This documentation falls into three general categories:

Test development aids/guidance (1) General performance test development guidance for developers, content experts, and QC personnel. This includes, but is not limited to, test/test item development.

(a) Job aids.

(b) Checklists.

(c) Local/major subordinate command (MSC)/TRADOC procedural guidance (for example, TRADOC regulations, pamphlets, JAs, IMI Implementing Instructions, and Learning Object Design specifications).

(d) Local/MSC/TRADOC policy guidance (for example, TRADOC, MSC, Army regulations, and local regulations).

(e) "Model" test items.

(f) Software (see app D).

Policy and procedure guidance (2) Policy and procedural guidance related to the development of the CTP and the SEP for a course (see paras [5-3](#) through [5-7](#), above, for more guidance). This includes reference materials and existing guidance necessary for the development of course policy and procedures

Doctrine and other existing materials (3) Doctrine, existing instructional materials, and SAT documentation, which includes all job/task analysis documentation, LOs, doctrinal materials from which instructional content is developed, and any existing similar instructional material from other courses required.

6-4. Select/review/revise performance objectives.

Expectations from performance LOs a. Whether developing a written test, or a performance test, a critical task is deciding not what to test (that is, the TLOs), but how to best test the objective. Usually, the performance objective clearly indicates what performance is demonstrated, how it is judged as correctly performed, the conditions of performance, and number of successful iterations of performance required for mastery. (If the performance objective includes all these factors, the task of the performance test developer becomes very clear—duplicate in the test item, with as high fidelity (match) as possible, the performance (action) called for in the objective, under the conditions called for in the objective (or a sampling of conditions), and evaluate the performance, using the criteria in the standard of the objective.

Note: The instructional material developer’s task is also made clear, and is similar; that is, provide learning opportunities that duplicate the action and conditions of the objective, and bring the learner to the level of the standards of the objective.

Reality bites b. Unfortunately, when the TDT receives the LOs, if they are not specific enough, the team may not immediately begin the test development process. A lack of clarity may exist in the exact action the learner is expected to perform after training, the necessary/alternative conditions of performance, or the standards expected/possible within the training setting. Or, the objective may match the critical operational real world task, to such a degree, that training and testing under the conditions, or to the standard(s) stated, are impossible, or dangerous. In either case, use [appendix E](#) if the LOs are unclear/untrainable as stated, and coordinate with the LO writer for changes. A great deal of the design of a good performance test item is directly attributable to the quality of the LOs.

Note: Do not attempt to design test items from poor objectives; modify the objectives first, then design the corresponding test items.

6-5. Design performance test items.

Performance test item steps

a. To design performance test items:

(1) Determine whether product, process, or both product and process measurement is needed. (See [para 6-6](#), below.)

(2) Construct checklist for product and process measures. (See [para 6-7](#), below.)

(3) Determine scoring procedures. (See [para 6-8](#), below.)

Determine level of simulation required

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

Fidelity

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

Factors in determining level of fidelity

d. To determine level of fidelity, consider such factors as the time required to perform a given task, the type of equipment required, the ability to present the task in a uniform (standard) manner, risk of damage to equipment or personnel, and the ability to evaluate an individual's performance with a high degree of objectivity. These considerations impose realistic constraints on the level of fidelity of the performance test item. Often, compromises are in order. Instead of requiring performance of a complex task, decide to limit the test to one or two phases of the task, such as preparing only one slide of a biological specimen, but identifying a larger number of mounted specimens. At times, computer simulations are used as a compromise.

Computer simulation for testing performance

e. Computer simulation of psychomotor task performance has its advantages and disadvantages, although it is frequently used in lieu of the use of actual equipment. Its value is directly proportional to its ability to simulate the actual task. For example, if the task is performed on, or with the help of, a computer, the level of fidelity is almost perfect, and there is no reason to acquire actual equipment. For procedural equipment tasks, computer simulation is most appropriately used, to determine the learner's readiness to practice and perform tasks on actual equipment. Certification of competency should take place on actual equipment, if the equipment is not simulated to a very high degree of fidelity.

(1) Advantages of computer simulation of psychomotor task performance:

(a) Protects equipment and personnel from dangers inherent to using actual equipment.

(b) Electronically records actions/mistakes for review.

(c) Easily simulates numerous faults/situations.

(d) Incorporates aspects of games for motivational purposes.

(e) Scales objects to workable size.

(f) Can determine learner's readiness to work with real equipment.

(g) Can possibly allow release of equipment for operational purposes.

(2) Disadvantages of computer simulation of psychomotor task performance:

(a) Only provides a 2-dimensional representation of 3-dimensional equipment (unless holographic).

(b) Cannot train the feel of an action (no tactile response).

(c) Learner may not feel the same degree of urgency, or criticality, to perform correctly.

(d) Does not match scale of actual objects.

(e) Not able to observe/record attitudes.

(f) Cannot provide closest approximation to real performance.

(g) More costly to develop high-quality simulations (if actual equipment is available for use).

6-6. Determine product or process measurement.

Introduction

a. In a performance test, the learner actually performs a task, and is judged against predetermined criteria. A performance test may involve product measurement, process measurement, or both. Job Aid [350-70-5.6a](#) provides guidelines for constructing hands-on testing. In developing a test plan, determine if the objective requires measurement by:

(1) A tangible, measurable product that the action produces.

(2) An observable process alone, disregarding any product produced.

(3) Both a product and a process.

Product measurement

b. Product measurement is always appropriate if the objective specifies a product. When a product measure is required, incorporate it into the LO, and carry it over into the test items. [Job Aid 350-70-5.6b](#) provides an example of a performance test measuring a product. An example is: "Navigate from point A to point B." Use product measures when:

(1) The objective specifies a product.

(2) The product is measured as either presence or characteristics, such as voltage, length, etc.

(3) The procedure leading to the product can vary, without affecting the product.

Process measurement

c. Process measurement is indicated when the objective specifies a required sequence of performances that is observed, and the performance is as important as the product. Process measurement is also appropriate in cases where the product is not measurable for safety, or other constraining reasons. [Job Aid 350-70-5.6c](#) provides an example of a performance test measuring a process. Process measurement is appropriate when one or more of the following is true:

(1) Detailed diagnostic information is desired.

(2) There is no product at the end of the process, or it is not measurable.

(3) The end product is not necessarily the result of the process alone (that is, perform life saving steps), or is dependent upon the actions of people or processes beyond the control of the learner.

(4) The product always follows from the process, but high costs, risks, or other practical constraints prevent measurement of the product.

Use of both process and product measurement

d. Some situations may exist where both process and product measurements are appropriate for a given objective. Job Aid [350-70-5.6d](#) provides an example of a performance test measuring a process and product. Several examples of conditions that may call for both product and process measurement follow:

(1) Although the product is more important than the process(es) which lead to its completion, there are critical steps, which, if not properly performed, may cause damage to equipment, or injury to personnel.

(2) The process and product are of similar importance, but do not assume that the product meets criterion levels.

(3) Diagnostic information is needed.

(4) Having process, as well as product measures, ensures information is obtained as to why the product does not meet the criterion.

Scoring for processes and product measurement

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

Product and process measurement roles

f. Three types of tasks illustrate the relative roles of product and process measurement:

(1) Tasks where the product is the process. Relatively few tasks are this type. Drill and ceremonies, playing a musical instrument, and public speaking are examples.

(2) Tasks in which the product always follows from the process. These are tasks, such as fixed procedure tasks. If the process is correctly executed, the product follows. For example, following the correct process to pack a parachute results in a properly packed parachute (the product).

(3) Tasks in which the product may follow from the process. In a large number of tasks, the process appears correctly carried out, but the product not attained. Two reasons this happens is either the developer was unable to fully specify the necessary and sufficient steps in task performance, or did not accurately measure them. Rifle firing, for example, illustrates that there is no guarantee of acceptable marksmanship, even when following all procedures. In this case, process measurement would not adequately substitute for product measurement. Therefore, if any uncertainty exists, that using only a process measure may not adequately measure a student's ability to achieve the product or outcome of a task, add a product measure.

6-7. Preparing checklists for process and product measurement.

Use

a. A checklist is useful for rating ability to perform a specific set procedure. It is also a simple method of rating performance skills, when the purpose is to see if learners have reached a certain minimum level of performance. Figure 6-1 shows a portion of a checklist rating form an observer uses for instrument flying proficiency, to indicate whether the completion of each step is satisfactory or unsatisfactory.

Instrument Flying Checklist		
INSTRUCTIONS: If the performance is satisfactory, place a '+' sign in the space provided. If the performance is unsatisfactory, place a '-' sign in the space.		
Item No.	Task	Rating (+ or -)
1.	Maintains constant heading within 5 degrees of course.	
2.	Maintains constant altitude within 50 feet.	
3.	Can make a timed turn within 10 degrees of a new heading.	
4.	Can make a steep turn within 50 feet of altitude.	

Figure 6-1. Checklist rating example

Developing a checklist for a process

b. When preparing checklists for rating a process:

(1) Avoid lumping steps together. Example: If step #6 is not possibly accomplished without successful completion of the preceding steps, it may not require measurement of those steps. Begin with the most logical step. Be careful not to mask a critical step when applying this technique.

(2) Determine the time standard, or qualitative or quantitative standard, that applies. All or a portion of the elements may need evaluating in a process task.

(3) Do not use a process test simply because the reference document shows a sequence. Ensure the sequence is critical, in terms of safety or security, to personnel or equipment.

(4) Specify GO/NO GO criteria for each element measured.

(5) Sequence steps in order of evaluation.

(6) Keep performance steps statements short and precise.

Developing a checklist for a product

c. When preparing a checklist for scoring a product:

(1) Specify the characteristics of the end product, prior to writing the test. Express standards in terms of (not all inclusive):

(a) Shape.

(b) Tolerance.

(c) Strength.

(d) Tightness.

(e) Time.

(f) Location.

(g) Texture.

(h) Consistency.

(i) Function.

(j) Speed.

(2) Specify the product characteristics as performance elements.

(3) Specify GO/NO GO criteria for each performance element and overall end product.

(4) When preparing checklists for rating a combination process/product test, combine the appropriate construction rules from both test formats. Regardless of the type of test, ensure measurement of all key elements (those critical to accomplishment, safety, or security).

6-8. Determine scoring procedures for performance measurements/ tests. Criterion testing lends itself to many forms of scoring. The purpose of a CRT is to differentiate masters from nonmasters by comparing them to an absolute standard.

Types of test scoring procedures

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

(1) Assist scoring.

(2) Pass/fail scoring.

(3) Fixed point systems.

(4) Rating scales.

Assist versus noninterference scoring

b. In testing, learners generally proceed from the beginning to end of a test, without comment or action on the part of the tester (noninterference). This type of scoring is often used in tests which call for the completion of a series of steps, or require production of a prespecified product. However, some tests may require scoring each step in a process. Thus, at each step, the learner's performance is approved (scored GO) or assistance is provided (and scored NO GO) before proceeding. Assist scoring is employed for diagnostic reasons. Remedial training then focuses on missed steps, saving retraining time and expenses. Assist scoring may also furnish valuable clues to areas where improvement in instruction is needed. For example, a large number of errors, in step number 3 of a 6-step procedure, may indicate a need for improved instruction in that area.

Example of
assist
method

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

Pass-fail
scoring

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

(1) A trainee is given 10 minutes to detect and replace a defective transistor in a radio set. The trainee either does (pass) or does not (fail) get the unit operational within the allotted time.

(2) The assistant gunner on the M-102 Howitzer has the responsibility for setting the quadrant on the quadrant sight and firing the weapon. The required processes are:

(a) Turn the counter handle to the appropriate numerical reading.

(b) Raise or lower the tube, until the bubbles on the sight are level.

(c) Fire the gun by pulling the lanyard on command.

Since this task is precisely checked for accuracy, a passing score is assigned only if no errors are observed on any of the above items.

Fixed point scoring

e. Fixed point scoring is another type of CRT scoring.

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

(2) An alternate type of fixed point scoring uses “pass-fail” decisions on components of a task. For example, trainees are asked to overhaul a carburetor, and a point value assigned to different components of the task. Table 6-2 provides an example of a pass-fail table, using a point value system.

Table 6-2
Pass-fail point value table example

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

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

Use of a checklist for scoring

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

Types of scoring

g. Employ the type of scoring system appropriate for the objective.

(1) If the objective specifies an action or product, use a GO/NO GO scoring system (either the action occurs in the proper sequence, or it does not; either the product results, or it does not).

(2) If the objective specifies characteristics of a criterion-level product or action, a rating scale, or other form of point assignment, is indicated. Point assignments are made on an explicit, well-defined basis for each item.

(3) For rating scales, inter-rater reliability is high. Tie point assignments to criterion levels specified in the objective.

Cut-off points

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

Determining mastery cutoffs for performance tests

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

(1) Number of iterations necessary as proof of mastery.

(2) Number of different conditions under which testing must take place to assure mastery.

(3) The scores to achieve (time, tolerances).

TRADOC Pam 350-70-5

- Consensus j. The entire TDT should participate in the mastery standard determination, but the opinions of the SMEs, with advice from the test developer, should predominate the decisionmaking process. The SMEs must agree on exactly what evidence to collect to determine mastery. If an agreement is not reached, include a third SME to arrive at a consensus.
- Classifying mastery k. In making decisions about mastery standards, make the test design team acutely aware of the costs of—
- (1) Wrongly classifying a master as a nonmaster (called a false negative classification), or
 - (2) Wrongly classifying a nonmaster as a master (called a false positive).
- Setting cut-off scores by costs of misclassification l. If the costs for false negatives are relatively high (e.g., manpower needs are critical) then lowering the cut-off score is justifiable. If the costs of false positives are high, then cut-off scores must remain high. In most cases, for training wartime critical tasks, the costs to personnel and mission of a false positive (certifying nonmaster as master) are much more serious than certifying a master as a nonmaster. In the former case, certifying a nonmaster as a master may directly put personnel, equipment, or critical mission accomplishment at risk. In the latter case, the only result is likely just retraining to standard, or some delay in getting resources to the field. Therefore, for most performance TLOs, the mastery standard is set very high to avoid certifying a nonmaster as a master. In any case, when performance on critical tasks is tested, cut-off points are kept high enough to reflect the standards specified in the objectives for those tasks.

6-9. Writing performance test instructions.

- Instructions for test participants a. Once you are satisfied with procedures, directions, equipment, and scoring methods for a performance test, prepare detailed instructions to the test participants to formalize them. The primary requirement for test instructions is that they are complete, clear, and provided in writing. They should also include diagrams and pictures, as necessary, to fully describe the test environment to the instructor/administrator, the observer/evaluators, any actors, and the learner. Job Aid 350-70-5.6e provides instructions and a checklist for a sample performance test.

Directions to the instructor and/or test administrator—

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

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

Directions to the observers/evaluators

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

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

TRADOC Pam 350-70-5

Directions to actors d. In some performance test situations, actors may perform certain actions necessary for the presentation of the scenario or situation to the learner. Actors work directly from a written script the test designer provides; and do not vary from the script, during any iteration of the test situation (that is, from learner to learner). You may find rehearsals necessary.

Directions to the learner e. In very simple situations, give oral directions to the learners.

(1) For example, to test a musician's ability, provide a sheet of music, and ask them to play the piece. However, such informality opens the door to the introduction of elements that could create nonstandardized testing conditions. An instructor might give more detailed instructions to one individual than to another, or might inadvertently omit something important from the instructions.

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

- (a) Purpose of the test.
- (b) Time limits, if any.
- (c) Equipment provided.
- (d) Requirements the learner is expected to satisfy.
- (e) Special safety precautions.
- (f) Information on how the test is graded.

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

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

Directions for scoring

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

(1) Including as many as necessary, to ensure comprehensive coverage of the action indicated in the objective. Provide sufficient evidence that the learner can, or cannot, perform the action to the standard indicated in the objective.

(2) Too few checkpoints probably indicate that some elements were overlooked.

(3) On the other hand, too many checkpoints may suggest a failure to differentiate between critical and trivial elements.

(4) The use of too many checkpoints may impose an impossible burden on the raters. It requires them to watch for too many things at one time, and possibly miss the important factors, while trying to grade performance on minor matters. For this reason, be selective and critical when developing the rating form.

(5) Picking the items that are significant to successful performance. Choose items of a nature that are observed and judged with a high degree of objectivity.

(6) Certain points may require the observer to check more than one item. For example, using a voltmeter in a physics project may require checking to ensure proper connection to the unit, and that the learner read the meter correctly. However, in many situations, you may desire the learner to record dial settings and meter readings on a separate form, specifically keyed to the instructions.

Include all safety items as checkpoints

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

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

6-10. Validation of tests/test items.

Components a. The validation of performance tests has four primary components. Revision takes place where necessary, and usually before proceeding to the next phase. In roughly the order in which they are initially performed, the phases/components are:

- (1) Content validation.
- (2) Review of instructions.
- (3) Individual tryout of items.
- (4) Master/nonmaster tryouts.

Conduct content validation b. After initial drafting of each item and instructions, match the performance in the item requested of the learner during the test, with the performance found in the objective, to verify content of the item. Then, match the conditions under which the test takes place, with the conditions listed in the LO. Lastly, match the rating scale/checklist against the LO standard, to decide its reasonable use at determining mastery or nonmastery. Include active reviews by TD and SME personnel throughout the item development process, to adequately accomplish this step.

Review of instructions c. For each item, an SME, other than one who participated in its development, should review all administrative instructions (including instructions to the administrator, observer/evaluator, and actors) for clarity, understanding, and completeness. The SME can also review the learners’ instructions, but the individual tryout, using an actual naïve learner, provides more information. Make necessary changes before going to individual tryouts. Document review and results.

Individual tryout of instructions d. Use the following procedures for conducting individual tryouts:

- (1) Select administrator, observer/evaluator, and actors from the likely pool of personnel actually administering the item.

(2) Obtain an untrained volunteer learner who best matches the characteristics of the target audience. A trained, but untested volunteer is best, but difficult to acquire.

(3) Train/provide instructions to administrators/actors in exactly the same way as during actual course test. If participants receive written administrator instructions, without the opportunity to ask questions of the developer, then present the instructions to the volunteers in the same manner. Collect comments for changes.

(4) Have the volunteer administrator train the actors and evaluators (if part of administrator's responsibilities). Provide assistance in training, only if absolutely required, and make note of problems/comments.

(5) Ask administrators/evaluators/actors to administer the test item to the volunteer exactly IAW the instructions given. Collect comments from administrators and learners.

(6) Make changes, as necessary, before proceeding to master/nonmaster trials. If the content or instructions are changed, conduct rereview of content and instructions. If the changes are significant, repeat individual tryouts for the item(s).

Note: Subject multiple items simultaneously to this tryout, following all guidance above.

Master/non-
master
reliability
tryouts

e. Master/nonmaster reliability tryouts provide the first statistical indicator of the reliability of the test. It is based upon the assumption that true masters of the TLO will pass the test, and true nonmasters cannot pass the test. Follow these rules to try out multiple items at one time:

(1) Acquire at least 5 nonperformers (with critical characteristics of the target population) and 5 performers as volunteer learners. Accept self-assessment of mastery only if other evidence of mastery is NOT available.

(2) Administer item(s) to volunteers, following written administrative instructions precisely. The administrators should answer only allowed questions from the volunteer learners. Stop test if it is clear that nonmaster cannot perform (or when failure is assured). Especially watch nonmasters for safety failures. Keep notes of mistakes masters make.

(3) Compute discrimination index and apply rules. Use [JA 350-70-5.7f](#). See [table 6-3](#), below, for actions based upon index computed.

(4) Repeat, if substantial changes are made. Revalidate any changes, to correct deficiencies that caused nonmaster to pass, or master to fail.

(5) Document results and gain approval (if passed), and include as part of the audit trial.

Table 6-3
Conclusions and actions from master/nonmaster reliability trials

Computed Index Range	Conclusions	Actions
.50 - 1.0	Item likely discriminates.	<ul style="list-style-type: none"> Accept as Reliable. Determine why any masters failed, or non-masters passed. Review comments collected. Make necessary changes.
.00 - .49	<ul style="list-style-type: none"> Item may not discriminate. One or more nonmasters may have some mastery. One or more masters may not be true masters. TLO does not require training; can assume as entry-level skill. 	<ul style="list-style-type: none"> Review item for adequacy. Review performance with nonmasters who passed item. Are they true nonmasters? Review performance of and discuss with masters to determine why mistakes were made. Check job analysis and verify assumptions of entry-level skills of target population. Modify item, as necessary.
Below .00 (negative number)	<ul style="list-style-type: none"> Item does not discriminate. One or more nonmasters are master, and/or one or more masters is nonmaster. TLO does not require training; can assume as entry-level skill. 	<p>Must research cause and revise item for adequacy, if found deficient. In particular:</p> <ul style="list-style-type: none"> Review performance of, and discuss with nonmasters who passed item. Are they true nonmasters? How were they able to pass? Clues? Review performance of, and discuss with masters who failed, to determine why mistakes were made. Check job analysis and assumptions of entry-level skills of target population. Modify/eliminate item, as necessary. Repeat validation, starting with content (content, instructions, individual, and master/nonmaster) for revised items.

Actions based upon computed index

f. Based upon the index computed, draw certain conclusions, and take action, IAW table 6-3.

g. Once the four processes above (that is, collecting documentation, selecting/reviewing/revising performance objectives, designing performance test items, and writing performance test instructions) are completed (and repeated, as necessary) for a set of test items, and necessary changes made, approve the test/items as sufficient for use during the instructional materials validation procedures. Approve the tests/items, based upon the results of these validation procedures, before using them to validate the instructional materials. See [TRADOC Pam 350-70-10](#), paragraph 1-6c, for more details.

Continuous improvement cycle

h. However, when administering the tests/items to a group of learners/volunteers, additional data regarding their quality is generated. Collect and use this data for the continuous improvement of the items. The individual, group, and operational trials during material validation, and each administration of the items to learners during an actual course of instruction, provide the opportunity to collect additional data. Correct identified and validated problems ASAP, before further use of the items. This data usually takes two forms:

(1) Encourage test takers and administrators to submit test critique comments.

(2) Collect learners' performance (answers, mistakes) on each test item.

Note: See paragraphs [5-10](#) and [5-11](#), above, for general responsibilities regarding data collection and analysis.

Analyzing performance test data

i. The administration of the performance test to a large group of learners provides the opportunity to conduct additional performance test item analysis. For example:

(1) If a large number of trained learners consistently miss an item, subject that item to review for deficiencies. Once the item (and its administration) is deemed sufficient, review the instruction for adequacy.

(2) Similarly, if a number of learners pass the item on a pretest without instruction, relook the test for clues or administration correctness. Again, once the item and its administration are deemed adequate, the training developer should reexamine assumptions about the entry-level knowledge of the learners.

6-11. Update CTP and SEP.

CTP and LAP a. Based upon the items developed, update the CTP with the information about the performance tests developed. Details are entered into the CTP, and may contain sensitive test information. Update the SEP with only the general information about the test of interest to the learners.

Sensitive nature of performance test items b. In most instances, the performance test item itself and the checklist is NOT sensitive in nature (for example, Put on Mission-Oriented Protective Posture equipment). In that case, it is provided to the learners via the SEP, unless each learner receives a different version of the test (for example, different indicators for different fault isolation and equipment fixes). See [chapter 8](#), below, for additional discussion on determining sensitivity of test/items.

6-12. Quality control criteria for developing performance measurements/tests. Quality control mechanisms are built into the steps above. In addition, the entire team should discuss the results of each step in the development and validation process. Documentation of procedures, results, conclusions, and action assures high-quality performance tests.

**Chapter 7
Development of Knowledge-Based Tests**

7-1. Knowledge-based tests overview.

Overview a. This chapter provides the details of the procedures to develop knowledge-based test instruments. Make and implement specific knowledge-based test development decisions (IAW the guidelines in chaps [3](#) and [4](#), above) during the writing of knowledge-based items.

Formats and knowledge-based tests b. A knowledge-based test seeks to measure/determine cognitive knowledge (that is, what a person knows), not (directly) what they can do. As stated in chapter 6, above, performance tests use “knowledge-based” test formats to solicit the learners’ response to a performance item. In almost every instance, the rules for developing and validating these formats are the same, whether they are used to measure knowledge, or cognitive skills (performances). Therefore, each performance test/item is developed and validated in one of two ways:

(1) As a pure performance test, using the checklists, evaluators, and techniques in chapter 6, above.

(2) As a pure cognitive skills, measurable by a written (knowledge-based) format, using the techniques in this chapter.

Note: The TLO determines the method of testing.

Knowledge-based test items and test instruments

c. When discussing knowledge/cognitive testing, a “test” is made up of several test items, which may use one/many of the acceptable test item formats. For TLO mastery purposes, a “test” consists of all test items used to measure mastery of a single objective. For differentiation purposes, a “test instrument” may have one or more knowledge-based TLO tests, administered at the same time (that is, a single written “test instrument” may contain several tests, with each test measuring learner mastery on a different TLO). For administration and mastery determination purposes, identify and evaluate the test items associated with each TLO, independently of test items for any other TLO.

Similarity to performance test development

d. Many steps and procedures for the development of knowledge-based tests/test formats are similar, or identical, to those for performance tests. When possible, this chapter refers to the preceding chapter to elaborate on the similar steps.

7-2. Knowledge test development steps.

General overview of knowledge test development

a. As with performance tests, in order to develop good knowledge/cognitive skills test items, ensure that the objectives are sufficiently specific, determine optimum formats, determine scoring procedures, write the items and instructions, validate items and instructions, obtain approval, and update the course test plan and the LAP. Again, note that the above steps are not necessarily linear in their performance—use spiral development for necessary corrections and improvements.

Review of key elements of performance items

b. As noted in [chapter 3](#), above, knowledge/cognitive skills test/items—

(1) Do not seek to, nor can they directly measure, noncognitive skills or task competency.

(2) Are used to test knowledge/cognitive skills that enable performance of the desired task/action (when it is not feasible to directly test the performance). From that information, make a prediction as to whether the learner is able to perform the task.

(3) Can determine if the learner obtained certain prerequisite knowledge/cognitive skills (defined during task analysis) necessary before actual performance is efficiently and effectively (and, sometimes safely) taught.

(4) Are useful, to the extent that they:

(a) Predict learner performance on actual performance tests.

(b) Measure knowledge/cognitive skills proven necessary for task performance.

(5) Can take the format of essay, short answer, fill-in-the-blank, labeling (or a computer-based version of labeling, called drag-and-drop), multiple-choice, or matching.

(6) Require learners to demonstrate mastery of supporting knowledge/cognitive skills, by responding to various types of written, oral, or computer-generated questions.

(7) Emphasizes intellectual knowledge/cognitive skills related to a performance objective.

(8) May require learners to find, read, and use technical materials.

(9) Are usually independent questions, and the test item sequence usually will not affect the outcome of the test (with possible exception).

Knowledge/
cognitive
skills test
development

c. Knowledge/cognitive skills test development encompasses the following steps:

(1) Collect documentation (same as for performance tests; see [para 6-3](#), above).

(2) Select/review/revise knowledge/cognitive skills objectives (see [para 6-4](#), above, and [para 7-3](#), below).

(3) Design knowledge/cognitive skills test.

(4) Write items and instructions.

(5) Revise and repeat above, as necessary.

(6) Update CTP and SEP.

7-3. Review and revise objectives.

Determining objective uses a. While reviewing all objectives, determine those that are clearly cognitive (knowledge-based); those that are pure psychomotor performance, and require evaluating via the assessment of a product or process; and those that are performance (cognitive skills), but are administered using a knowledge-based format, to collect and evaluate the learner's responses. The first and the last of these types are set aside for the development of tests using knowledge-based formats. For reference purposes in this chapter, the term "knowledge/cognitive skills" is used to describe these types of test items.

Knowledge/cognitive skill objectives b. Knowledge/cognitive skills objectives frequently suffer from the same types of deficiencies as performance objectives. That is, they are not specific enough to immediately begin the test development process, and require revision before proceeding.

Rationale for lack of clarity in knowledge/cognitive skills c. In knowledge/cognitive skills objectives, as for performance objectives, there may exist a lack of clarity because—

(1) The exact action the learner is expected to perform is unclear, usually because of overuse of the verb in the knowledge/cognitive skills objective (for example, "describe"), when another verb might make the objective clearer.

(2) Certain conditions are lacking—most notably a clear "with" or "without references/notes/job aids" condition and the specification of those references/aids.

(3) The standard is not specific (for example, standards frequently imply or state "without error," and the real-world standard is less than perfect (and/or the cutoff scores are usually not set at a "perfect" level).

Note: Use [appendix E](#) if the LOs are unclear. Do not attempt to design test items from poor objectives; modify the objectives first, then design the corresponding test items.

7-4. Design knowledge/cognitive skills items.

Steps in design a. The following steps cover the design of knowledge/cognitive skills items:

(1) Determine if you will use recall or recognition-type items.

- (2) Select the best format for objective and required item type.
- (3) Determine number to adequately measure TLO/ELOs.

Step 1:
Determine
recall or
recognition

b. An often-overlooked step in the design of knowledge/cognitive skills items is to determine if the objective calls for the recall of information, or just the recognition. Recall objectives assume that the learner must “recall, and state from memory, without the presence of the correct answer” certain facts, procedures, policies, steps, etc. Recall formats do not provide the learner alternatives from which to choose the correct answer. On the other hand, recognition means only that the learner “recognizes” the correct answer when they see it. In a recognition format test item, the learner is given a relatively small set of alternatives, and must select the correct answer from this set.

Note: An objective requiring the learner to identify something/ someone, etc, is not necessarily a “recognition” item. The manner in which the learner delivers the answer determines recall or recognition. If the learner identifies and picks the correct name from a list (or drag-and-drop the correct name), then the item is recognition. If the learner identifies, by writing, the name—without options to choose from—then it is a recall item.

Types of
recognition
and recall
formats
Why recall is
better

c. Recall type items include essay, short answer (or long answer), completion, including fill-in-the-blank, and labeling (not drag-and-drop). Recognition formats are drag-and-drop, matching, and multiple-choice.

d. Recall test items are the better measure of true knowledge/cognitive skills acquisition and mental ability, since no hints (the correct answer) are present in the item. The answer is retrieved from memory. Recall formats are better for three other reasons.

(1) Recall-type items eliminate the possibility of guessing the correct answer.

(2) Recall items (correctly answered) provide better retention.

(3) Good recognition-type items are, by their nature, more difficult to construct. For example, it is more difficult to construct a good multiple-choice question with no design flaws, and the required three to four reasonable distracters (incorrect answers), than to ask the question directly.

Recall formats and performance tests

e. Recall-type formats are more valid measures, whether they are used to solicit knowledge, or cognitive skills (performances). A cognitive skill item, that asks a learner to demonstrate skill by writing in an answer, is usually better than one that requires selection of the correct answer from a number of choices. The learner may know absolutely nothing about how to solve the problem, but may eliminate (because of construction problems) several given alternatives, and then have a good chance of correctly guessing the answer.

Guidance on use of recall versus recognition

f. Use recall or recognition IAW the following:

(1) Use recall items for all TLOs (knowledge/cognitive skills) that are absolutely critical for safe task performance and critical wartime mission accomplishment. This requires a review of the job and task analysis information, to see which performance(s) (tasks) the knowledge/cognitive skill supports, and/or the performance objectives supported. Recall items are recommended for all other objectives.

(2) Use recognition items if, and only if, the test item is relatively noncritical, and a conscious decision is made and approved that the characteristics of recall objectives sufficiently outweigh their decrease in measurement value.

(3) Do not use recognition items solely because of their perceived ease of grading. Due to their format, the primary “advantage” of recognition items is their perceived ease of grading. However, with state-of-the-art word recognition and matching software, this advantage is quickly diminishing.

Step 2: Determine specific type of items

g. If testing recognition is sufficient, construct multiple-choice (also called “multiple-guess”), matching, or drag-and-drop (another type of matching.) If the objective needs recall, use essay, completion, short answer/fill-in-the-blank/labeling. Select the type of item based upon the guidance below. The use, advantages/disadvantages, and construction guidance of each type is covered in [paragraph 7-5](#), below.

Step 3: Determine number of items

h. Use the following guidance to determine the correct number of items to measure a cognitive skill or knowledge:

(1) Usually one test item can measure acquisition of one piece of knowledge. (Matching items might measure several pieces in one item.)

(2) If the objective calls for only one piece of knowledge (for example, “write Ohm’s Law”) then one item might measure it adequately, although you should develop multiple versions of the item for pretests and alternative version of the posttest.

Note: If testing several ELOs, you probably need at least one item (with several versions) for each ELO.

(3) Measuring whether the learner knows a series of steps may require multiple items, to test knowledge of all steps (that is, if each step requires different skills/knowledge).

(4) For critical cognitive skills, using one item may not suffice to verify mastery. The SME expertise determines how many items/different trials to use. For example, an objective like “determine the number of matching points for two fingerprints,” requires multiple trials.

Note: See guidance in paragraph [3-9c\(2\)](#), above, for additional information.

7-5. Use, selection, and construction of knowledge-based test items.

Strengths
and
weaknesses

a. Each type of test item has certain strengths and weaknesses, in regard to their use and ease of construction. This paragraph describes the common types of items. For general guidelines for the construction of all types of knowledge-based test items, see [JA 350-70-5.7a](#).

b. Objective knowledge-based written tests offer several advantages:

Advantages
of
knowledge-
based
written tests

- (1) They are reliably administered.
- (2) They are machine-scored.
- (3) They cover a large amount of material, in a short period of time.
- (4) Test score data is easily maintained for recordkeeping purposes.
- (5) Statistical data, describing certain test item characteristics, such as difficulty, the mean and variance of test items, correlation between test items, response patterns, internal test consistency, and test variance, are computed.

Disadvan-
tages

c. Disadvantages of the items are discussed under each type item below, and in paragraph [7-6](#), below.

Multiple
choice test
items

d. The multiple-choice test is probably the most widely used test item. It is composed of a stem, an answer, and distracter. It is written in the correct answer method, or the best answer method.

(1) The multiple choice test is used to measure simple, and somewhat complex, LOs. These include:

- (a) Principles.
- (b) Methods and procedures.
- (c) Specific facts.
- (d) Applying facts and principles.
- (e) Interpreting cause and effect relationships.
- (f) Justifying methods and procedures.

(2) The main advantage of multiple-choice tests is wide applicability. The use of distracters dispels common misunderstandings that need correcting, and the multiple choice test is also easy to construct and score.

(3) Their disadvantages include:

(a) Locating good distracters. (At least three are recommended. Use two distracters only if three feasible alternatives are not possible.)

(b) The correct answer is given, instead of student-provided (that is, it may measure only recognition, not recall).

(c) Test item construction errors can frequently “give-away” the correct answer.

(d) The probability of guessing a four-choice item is 25 percent. As implausible distracters, or poorly designed alternatives, are eliminated, the probability is quickly increased to 50 percent, which is no better than a true-false item. A lengthy list of “Dos and Don’ts” for multiple-choice test item construction is found in [JA 350-70-5.7b](#). Use this list to check your completed items.

TRADOC Pam 350-70-5

Matching test item

e. When the objective requires the learner to make comparisons, or identify relationships between two sets of items, use the test item construction of the matching test item. General guidelines are:

(1) Place the “premises” in the left column, written to the necessary detail, and of one category, or homogeneous.

(2) Place the responses in the right column.

(3) Write items so that there are at least two more logical, possible (but wrong) responses (possible answers) than there are premises. This avoids the learner arriving at the last answer through process of elimination.

(4) Also, to help avoid this process of elimination, instructions should state that each response is used one or more times, or not at all.

(5) Directions are usually required for the test item. They inform the student on what basis the association is made.

(6) Use this test item mainly for measuring simple knowledge/cognitive skills. The student relates terms to definitions, controls and functions, and so on.

(7) The main advantage to matching test items is testing a large amount of information at a time.

(8) The main disadvantage is difficulty in development, to avoid irrelevant clues. Also, as stated, it is usually used to test simple knowledge/cognitive skills, such as recall, not real understanding.

Drag-and-drop variation of matching item

f. Another variation of the matching format is used for computer-delivered test items. Using the mouse, or other input device, the learner drags the correct answer (either the words themselves, or the letter/number designator of the correct answer) from a list of possible answers (the response), to the place on the screen where it matches the premise. In this case, the matched item (premise) is either a picture (labeling), or just words/symbols/etc., like a written two-column matching item. Table 7-1 provides a checklist for developing matching test items.

Note: Use a drag-and-drop technique to place things in their proper place on a diagram or picture (that is, hook up cables, attach appendages to equipment, etc.) Frequently, this use of drag-and-drop becomes a low-fidelity simulation of actual actions (steps) of a performance objective, and is, therefore, a performance test of the

cognitive skill component of a psychomotor skill/task, without the motor (tactile) component of the real equipment. In this case, the rules for development of performance items (that is, simulations) apply.

Table 7-1
Checklist for developing matching test items

See JA 350-70-5.7c for more guidance to use as a checklist for matching test item development and review.	
1.	Are premises all of one type (homogenous)?
2.	Is the number of premises 10 or less?
3.	Are the instructions stated so that responses are one time, more than one time, or not used at all?
4.	Are the responses briefly stated, with the premises written to the necessary detail?
5.	Are the responses written in logical order?
6.	Are all test item premises and responses on the same page?
7.	Is the basis for comparison between premises and responses clear or clearly stated?
8.	Do premises and responses match grammatically?

Use of completion test items

g. In general, if the essay type question is not required, the best type of item to measure a knowledge/cognitive skill is the completion. Completion test items are:

- (1) Useful for testing an item of knowledge, in a specified context.
- (2) Appropriate for testing intellectual skills, such as discrimination, concrete concepts, defined concepts, and rule learning.
- (3) Appropriate for testing associations, some discriminations, portions of chains (fill in the missing steps) and intellectual skills, related to more complex types of behavior, such as declarative knowledge (verbal information).

Short answer/completion test items

h. The short-answer/completion item is written as a complete question (recommended), or as a fill-in-the-blank (a poor second choice). It has three advantages: ease of construction; the answer is student provided (that is, a recall item); and reduces guessing. A disadvantage is difficulty in wording, to obtain the desired response (especially for fill-in-the-blank items). Therefore, write a complete question, if at all possible. Completion items are somewhat more difficult to machine-score than matching or multiple choice tests. To construct short-answer/completion test items, use the following guidelines (also, see the checklist in [JA 350-70-5.7d](#)):

TRADOC Pam 350-70-5

Construction of short-answer/completion test items

- (1) Word items for brief, direct answers.
- (2) Do not take sentences directly from the lesson text.
- (3) Use direct questions, when possible.
- (4) Provide units of measurement, when possible.
- (5) For completion items, place blanks for answers to the right, all in equal lengths.
- (6) For fill-in items, do not use too many blanks.

Essay test items

i. The essay test item is normally a less used type of item, and requires the learner to provide a lengthy written response to instructions or a question. One advantage of the essay is that it tests complex knowledge/cognitive skills, which no other form can test. However, essay test items are difficult to develop and score objectively. Essay test items are written in two forms—restricted response limits the student's response as specified in the directions; and extended response allows almost total freedom; but, it is more difficult to score. The uses of the essay test item include measuring the student's ability to organize, integrate, and relate facts and principles, application of those principles or facts, and similar measures.

Note: Making a list in response to test instructions or a question is thought of as a long-answer item, or an essay item. In this case, review both sets of guidelines below for its construction and use. (See additional guidance at [JA 350-70-5.7e.](#))

Develop essay test items

- (1) To develop essay test items:
 - (a) Do not use essay tests when you can use more objective forms.
 - (b) Write test items for the specific subobjective.
 - (c) State each test item so that the student's task is clearly understood.
 - (d) Indicate appropriate time limit.
 - (e) Avoid use of optional questions.

Scoring
essay test
items

- (2) To score essay test items:
 - (a) Prepare an outline of acceptable answers before grading.
 - (b) Use the appropriate evaluation method.
 - (c) Decide how to handle irrelevant factors.
 - (d) Evaluate all answers to each question at the same time.
 - (e) Evaluate answers, not students.
 - (f) Use independent evaluations for important tests.

Write
scenarios

j. For most knowledge-based tests/test questions, scenarios to “set-up” the question are not necessary. Ask for the knowledge, by the most direct means, without unnecessary elaboration. However, if necessary, use a scenario to describe the essential set of conditions and assumptions that the learner needs to consider/use for formulating the answers (usually for a grouped set of related questions, rather than for a single question).

Write
instructions
and scoring
key

k. The last step in the construction of knowledge-based items—prior to formal test validation procedures—is to construct the scoring key and write all necessary test instructions. These instructions include:

(1) Test administrator instructions, including those necessary for proctors at remote locations who are not instructors/SMEs (that is, for DL courseware).

(2) Learner test Instructions for:

- (a) The entire test.
- (b) Sections of test having different type items.
- (c) Individual items requiring instructions.

(3) Test response evaluation (grading) instructions and TLO GO/NO GO criteria.

Scoring key
and passing
(cut-off)
score

l. For each item, develop a scoring key, listing the correct answer (or, for essay questions, the criteria for awarding GO/NO GO) for the item. Also, list the GO/NO GO criteria for each TLO (the number of correctly answered questions required to receive a GO for that objective). Also,

update specific actions of the administrator/others, unique to each test, and include as a part of the administrator instructions. Refer to [appendix B](#) for more detailed guidance.

7-6. Validating knowledge-based test items. The validity of knowledge-based written tests is a major problem. Even if LOs relate directly to job performance, it is often difficult to relate knowledge-based written tests directly to job performance. Consider these examples for different types of test items:

Multiple-choice test items

a. Multiple-choice test items require the student to discriminate between several possible answers to select the correct one (recognition test item). Validity problems occur because the only type of job objective, for which the multiple-choice test item is directly suited, is one that requires the student to select from among several alternatives on the job. Examples of selection from a set of alternatives on the job are:

(1) Selection of tools for a given purpose.

(2) Selection of a proper procedure, from several described in a technical manual.

Matching test items

b. Even though matching items might appear appropriate for testing associations between concepts, problems can arise with validity of the test items. For example, a student could match a list of telegraphic code dots and dashes with the appropriate letters. This is not a directly valid measure, since the job requires translating an audible code, not a visible one, into the message.

Completion test items

c. Validity problems are also possible with completion test items, especially if the direct relevance to job performance is questionable. For example, a completion requires a student to list the four major distinguishing features of quartz. The student may pass the item from memory, but that does not verify that the student can always identify quartz when given a variety of mineral samples.

7-7. Validate test items.

Test validation overview

a. This paragraph provides information about and guidance for the conduct of the primary steps/components in the validation of knowledge-based test items. Conclusive “validity” is a subjective concept that is assumed through the application of these test steps, and further approached through the continued improvement of measuring instruments. The three primary components of knowledge-based test item validation are:

(1) Construct review to ensure test items match the content and objectives (i. e., “content validity”).

(2) Construct review to ensure the application of best practices and avoid flaws in test item design (that is, “construct validity”).

(3) Discrimination analysis (also called indirect validity) to determine if an item can distinguish between masters and nonmasters.

Note: Some authors describe discrimination analysis and/or content review as measures of “reliability,” not “validity.” Nevertheless, the application of the procedures above will result in determination that the items have sufficient validity and reliability, for use in learner measurement and mastery determination.

Other sources of validity data

b. Another information source of data to improve test items is the comments from instructors and learners during any administration of the items (that is, validation trials and implementation), and the actual performance of the learners. This data is analyzed to determine any problems with the test items undiscovered through the above validation processes. Use [table 7-2](#), below, to collect comments.

Content validity definition

c. Content validity is a measure of how closely the test instrument relates to the content of the instructional program it is designed to measure.

(1) Directly associate both the test questions and the course content with specific instructional objectives.

(2) Use formal task analysis methodology to systematically sample the content domain.

(3) The content domain tasks and subtasks identified through the task analysis process should form the basis for instructional objectives.

(4) The degree of content validity is not expressed numerically. It is described in terms of the comparison between, or the correspondence among, course objectives, course content, and test questions.

(5) Evidence of the degree to which a test is deemed as content valid is based on the combination of the following:

(a) The comparison of test questions with course content.

(b) The comparison of test questions and course content with instructional objectives.

TRADOC Pam 350-70-5

Update CTP and SEP

d. As a final step, document the test items, test item groupings, test groupings (coverage by TLO), mastery scores, administrator test instructions, learner test instructions, test proctor instructions, etc., in the CTP, and include appropriate excerpts in the SEP. For audit trail purposes, maintain the rationale for decisions made and approvals acquired in the test development documentation.

Construct validity

e. Construct validity is defined as the degree of application of proper test design procedures, to avoid test items that are confusing, unclear, biased, or have design flaws which assist in the determination of the correct answer. Construct validity is not an absolute, but skilled test developers carefully reviewing the items will remove most of the design flaws, thereby assuring an acceptable degree of construct validity.

Construct review procedures

f. The steps in conducting a construct review somewhat mirror those of the content review:

(1) Select, as a minimum, one skilled test designer (other than the original test item writer-developer) to perform the review.

(2) Conduct the review of each type of test item using the JAs mentioned previously in this chapter as checklists (JA 350-70-5-7a-e).

(3) Provide feedback from this review to the test developer.

(4) After test item changes are made, repeat the initial content review process (steps 1-3) for any test items that underwent major revision.

(5) Maintain completed checklists for audit trail.

Discrimination analysis

g. One way of "validating" test items is to look at their ability to discriminate between performers (masters) and nonperformers (nonmasters). A simple method of determining the discriminatory power of a test item, or test item set, is to administer the item/set to a sample group, composed of individuals who are clear (proven) performers of the objective to test, and another group composed of individuals who are clear nonmasters. Compare the performance of the two groups, to determine whether the masters consistently answered the question correctly, while the nonmasters consistently did not answer correctly.

(1) Use a statistic (the phi coefficient) to compare the two groups. The coefficient can range from a -1.0 to a +1.0.

Note: Use the excel spreadsheet in [JA 350-70-5.7f](#) to perform the computations.

(2) If the coefficient computed exceeds a certain predetermined threshold level, the item is accepted as an adequate discriminator. If the coefficient computed is lower than a predetermined threshold, the item is tentatively rejected, and requires review and/or revision (and subjected again to content, construct, and discrimination analysis). Table 7-2 lists recommended coefficient thresholds and actions.

Table 7-2
Thresholds and actions from master/nonmaster test item analysis

Computed Index Range	Conclusions	Actions
.50-1.0	Item likely discriminates.	<ul style="list-style-type: none"> • Accept as reliable. • Review comments collected during administration. • Make any necessary changes.
.00-.49	<ul style="list-style-type: none"> • Item may not discriminate. • TLO does not require training, can assume as entry-level skill. 	<ul style="list-style-type: none"> • Review item for construction flaws. • If several nonmasters pass, review entry level skill assumptions of target audience. • Modify item as necessary.
Below .00 (negative number)	<ul style="list-style-type: none"> • Item does not discriminate. • One or more nonmasters is master, and/or one or more masters is nonmaster. • TLO does not require training, can assume as entry-level skill. 	<ul style="list-style-type: none"> • Must research cause, and revise item for adequacy if found deficient. In particular: <ul style="list-style-type: none"> ◦ Review performance of, and discuss with nonmasters who passed item. Are they true nonmasters? How were they able to pass? Clues? ◦ Review performance of, and discuss with masters who failed to determine why mistakes were made. ◦ Check job analysis and assumptions of entry-level skills of target population. • Modify/eliminate item, as necessary. • Repeat validation, starting with content (content, construct, and master/nonmaster) for revised/new items.

Collect data during material validation

h. After determining each test item is acceptable, use it for the conduct of individual and group trials of instructional materials. During the instructional material trials, immediately correct any problems with any test item, in order to use it for instructional material assessment and

TRADOC Pam 350-70-5

validation. Based upon the type of test item, during these trials, collect additional data to use to assist in setting cutoff points.

Procedures for collecting data during validation trials

i. Use or modify the following common steps to assist in obtaining information to improve test items:

(1) Administer the test (consisting of the new test items) to the tryout sample.

(2) If the test is timed, record the time it takes each learner to complete the test, to assist in establishing the official time for the test. Obtain feedback from the tryout sample, by requesting identification and comment on any problem questions. During tryouts to provide feedback to improve quality of test items, test reviewers and learners may use the questions in table 7-3, below.

Table 7-3
Provide feedback

Question	Comments
1. Were any test instructions unclear or confusing?	
2. Was any question confusing?	
3. Did any question have more than one correct answer or no correct answer?	
4. Did any question have clues to the correct answer, either in that question, or in a previous question?	
5. Were any of the incorrect answer choices unrealistic?	
6. Did any question have unclear or confusing graphics?	
7. Did any question have language or graphics that were offensive, or contain bias or confusion related to race, gender, or cultural differences?	
8. Were any instructions on how to respond to each test item unclear or confusing? Modify item as necessary.	

(3) Perform appropriate criterion-referenced analysis on the data collected. Since the test is normally given to an untrained sample of volunteers during the trials, and then again after they are trained, it is appropriate to use the PHI coefficient again to identify weak items. (See procedures in paragraph [g](#), above.)

(4) Use the feedback from paragraphs (2) and (3), above, to decide which test items need revising or replacing. (The test developer makes

this determination.)

7-8. Compiling knowledge-based test items.

- | | |
|---------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Compile test instrument | a. If desired, consolidate mastery testing of several TLOs into a single administration. However, failures may result in the repeat of several of the lessons, based upon the sequential and progressive design of the instruction. |
| Grouping items | b. In compiling the test, test items of the same type are usually administered together. However, if testing multiple objectives in the single instrument, for convenience, group the related questions for each TLO. |
| Grouping affecting length | c. Follow guidance in paragraph 3-9c(4) , above, for the length of the test. |
| Update CTP and SEP | d. As a final step, document the type of items, number of items, test item groupings, test groupings (coverage by TLO), mastery scores, and other test design decisions in the CTP, and include appropriate excerpts in the SEP. |

7-9. Quality control criteria for developing knowledge-based test items.

To assure the development of effective knowledge-based tests, use the checklists provided, reviews of skilled test developers and SMEs, collect and analyze comments and learner performance data, and adhere to the procedures in this chapter.

**Chapter 8
Test Administration and Control**

8-1. Test administration. Guidance in the process of test administration and the procedures to employ to control sensitive test materials are discussed below.

- a. Test administration consists of the processes of—

- | | |
|---------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Reproduction and delivery | (1) Reproduction, as necessary of test material. Many tests, and their associated material, are digitized and placed in training product repositories, for control, access, and delivery. Use electronic storage and delivery of tests and related test materials (such as the LAP). Employee centralized storage, reproduction, and distribution of nondigitized testing materials. Determining factors for reproduction include: |
|---------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

- (a) The location and number of the administering activities.
 - (b) The number of learners at each activity.
 - (c) The delivery means (that is, hardcopy or electronic).
- (2) Delivery of test materials IAW existing SOP.
- (3) Securing of test materials during reproduction, distribution, and administration (see para [8-3](#), below).

Administering activity responsibility

(4) Administration of the test to the learners IAW the test instructions. It is critical that personnel administering the test instruments are fully aware of the proper control procedures, and the contents of the test administration instructions the proponent provides. The course administration documentation should include specific test control procedures to employ for each test. In particular, the test administering activity:

(a) Administers all test instruments exactly IAW TD proponent-provided guidance (or, if not possible, or if concerns arise, immediately contact the TD proponent for guidance).

(b) Implements the appropriate test control procedures, based upon guidance the proponent provided, and this chapter.

(c) Administers test reviews guidance the test proponent provided, and guidance contained in this pamphlet.

(d) Administers test critique instruments, and provides data to the TD proponent.

(e) Provides recommendations/concerns for test improvement to the TD proponent.

(f) Consults the TD proponent for questions regarding reclaims.

(g) Develops a clear and detailed local SOP for test control, IAW paragraph [8-3](#), below, and additional guidance from the proponent.

(5) Collection of learner performance data and comments.

(6) Delivery of learner performance data and comments to the proponent.

Documentation to assist with implementation

- b. Administer tests under standardized conditions. Ensure the test material includes accompanying documentation that specifies:
- (1) Test administration conditions.
 - (2) Instructions.
 - (3) Administration procedures (including how to handle questions, how to check and set up test supplies and equipment, etc.).
 - (4) Circumstances for excusing examinees from the test, due to illness, fatigue, etc.
 - (5) Environmental circumstances under which test administration is cancelled.
 - (6) Scoring procedures.

QC for test administering activities

- c. See responsibilities in paragraph [a\(4\)](#), above. Also, the administering activities' self-assessments of test control and administration procedures, and assistance/accreditation visits from TRADOC teams, provide QC during implementation. See [JA 350-70-5.8a](#) for additional information for administering tests.

8-2. Controlling testing material.

Purpose and applicability

- a. This paragraph provides guidance and procedures for the proper control of learner performance measuring instruments (tests), test items, and related sensitive material, such as specific scenarios and scoring keys.

(1) Although primarily intended for active and reserve component test administrators, and other test instrument handlers at the testing site/institution, it also provides information and requirements critical to test proponents, reproduction and transmittal activities, and any other activities/personnel that handle sensitive test materials.

(2) If needed for administrative ease or standardization, and for reduction in test compromises, local test administering activities may establish additional internal test control procedures beyond those required in this pamphlet. All developing activities should use care to ensure the procedures are of significant value in reducing test compromises, without unnecessary administrative burden, delay, or cost to personnel, or other training/education functions.

TRADOC Pam 350-70-5

(3) Handle security of classified tests IAW the appropriate regulations (that is, [AR 380-5](#)). (This subject is not further addressed in this pamphlet.)

Definitions

b. See the glossary for definitions for [test control](#), [test](#), and [sensitive test material](#).

General responsibilities of personnel

c. Personnel handling, or coming in contact with, sensitive test materials are responsible for their security.

(1) The learner is primarily responsible, logically and legally, for ensuring inappropriate disclosure/acquisition does not occur. Direct learners to report/identify all possible test material handling situations that might lead to inadvertent test compromise.

(2) All personnel, who may intentionally or unintentionally come in contact with sensitive test materials, are responsible for reducing the possibility of unintentional disclosure of test items or materials (test compromise). In particular, the staff and faculty of the test administering activity have a primary role in implementing these procedures.

(3) All commanders, staffs, department/division heads, instructors, and other personnel who might come in contact with sensitive test materials are responsible for limiting test material access to those individuals with an absolute “need to know” status. As with classified material, rank or position is not the primary deciding factor in determining “need to know.”

(4) Commanders/commandants, and training/TD (task) proponents are responsible for implementing the appropriate level of test and test item control. This requirement applies to all activities with sensitive test material under their control that requires securing. See subparagraph [e](#) below, and the remainder of this paragraph, for a full explanation of what this means for the test administering organization.

Basis for test control policy and procedures

d. The Army’s rationale for the determination and application of test controls is summarized as follows:

(1) Ensure test control policies and procedures, selected and implemented, apply appropriate measures, to reduce the possibility of test compromise to an acceptable level, without unduly burdening personnel or functions.

(2) The need for elaborate test control procedures decreased with the move towards more performance testing, and embedding knowledge-based tests into electronically delivered courseware.

(3) Although the possible seriousness of test compromise is potentially extreme (that is, certifying a nonperformer as actually knowing how to perform); normally, before life, mission, or equipment is put at risk, the former test-taker is further evaluated, coached, and/or observed. This ensures the certification obtained in institutional training is accurate and transferable to the job, or to later learning activities.

(4) All staff and faculty members of the administering activity are in a trusted position, and expected to exercise due caution and common sense in the handling of sensitive test materials.

e. For comprehensiveness, the applicable regulatory guidance pertaining to test control, found in enclosure 4 of [Memorandum, HQ TRADOC, ATTG-CD, 6 Oct 01, subject: Student Academic Measurement/Testing Policy Guidance](#), is reiterated below. To accomplish the implementation of the appropriate level of test and test item control required, administering organizations should:

(1) Maintain security of all test items, tests, test administration instructions (if necessary), checklists, scoring keys, and test results during test development, transmittal, storage, retrieval, and administration, consistent with the appropriate level of test control, as determined by applying the guidelines in the following paragraphs. See [JA 350-70-5.8c](#) and [JA 350-70-5.8d](#).

(2) Develop and specify in a lesson plan, and test SOP (if desired), the exact administration procedures to follow during resident test administration, to ensure the proper level of test control.

(3) Regardless of how final test/test items are ultimately administered, restrict access to paper-based copies of proposed or final test items, scoring/answer keys, or test results, to those personnel demonstrating a valid need for the information.

(4) As necessary, in conjunction with information management specialists, develop and specify procedures, to ensure electronic copies of tests/test items and scoring/answer keys are protected from unauthorized disclosure. These procedures:

Must include ---	May include ---
Restrictions on access, reproduction, and distribution.	Authentication methods.
Password protection.	Encryption technologies
Required learner warnings/certifications.	System intrusion detection/prevention methods.

(5) Immediately investigate suspected compromises, and take appropriate actions to reduce the impact of test/test item compromises.

(6) Ensure the test is administered exactly IAW the test administration instructions.

(7) Ensure test control procedures include a method to determine, with assurance, the identity of the test taker.

f. As found in enclosure 4 of Memorandum, HQ TRADOC, ATTG-CD, 6 Oct 01, subject: Student Academic Measurement/Testing Policy Guidance, use the following general guidelines for the control of tests:

Guidelines
for control of
tests

(1) Test items that mirror the task/TLO, and always performed in exactly the same way, require no security measures for adequate test control. For example: "Perform CPR," "disassemble an M-16A1," etc.

(2) Test/test items, with only a few possible variations, require that all possible variations are protected from unauthorized disclosure.

(3) Every variation of tests or test items require sufficient security measures, to avoid making known to the individual learner the specific variation they receive (that is, due to its size, making known the entire domain of possible test items is not detrimental to learner testing). For example, for testing purposes, a learner is required to repair 5 of 30 possible commonly occurring faults in a microcomputer. Making known to the learner the entire domain of 30 faults is unimportant, but once a particular subset of 5 faults is assigned for learner evaluation, do not make the subset known to the learner until appropriate.

Transmitting
and labeling
sensitive
material

g. Ensure the test proponent designates adequate procedures for test security when mailing sensitive test material. If the following controls are not applied to the tests received, apply them immediately (if possible) and/or contact the test development/reproduction activity. The test development/reproduction activities should apply the following controls:

(1) Clearly label each page of all sensitive test material requiring control (that is, except for Type 1, IAW para f, above) with "Examination (Testing) Materials – Sensitive in Nature," or similar words, to clearly indicate their nature. This includes paper copies or portable disks/diskettes (floppies/.zip files/compact disk (CD)). Label paper files containing test materials with similar markings, as well. Note: "For Official Use Only" is not appropriate.

(2) Ensure the first page of all controlled testing material, whether paper or electronic in nature (when displayed on the screen), includes the label indicated in paragraph (1), above. It is highly recommended that paper versions of tests include the warning on each page, as well as the cover sheet.

(3) All electronic versions of sensitive test materials should include the warning in paragraph (1), above, as well as the warning “DO NOT COPY, PRINT, TRANSMIT, OR SAVE UNLESS SPECIFICALLY AUTHORIZED,” or words to this effect, on any portable medium, and on the first page/screen when the file is opened.

Common controls

h. The administering activity should implement the following common controls for all test materials and test administration situations:

(1) Positively verify every learner’s identity before test administration.

(2) Before the administration of the first test in a course (usually at the beginning of each course/separate phase), advise all learners of the following:

(a) They will not acquire or provide inappropriate assistance before, during, or after any test, except as instructed (that is, for group activities).

(b) They will report any unauthorized assistance (before, during, or after the test administration) of which they are knowledgeable.

Note: Provide first-time notification in writing (that is, during course inprocessing), and obtain learner acknowledgment. Learner signature on acknowledgment is highly recommended. Also recommend reiteration of the above before each test.

(3) When not actually in use, ensure all sensitive testing material is under the appropriate control, as indicated in paragraph [j](#), below.

Additional controls based upon type of sensitive material

i. In addition to the common controls above, apply the following controls, as indicated, based upon the nature of the sensitive material.

(1) Hands-on test that mirrors task: No controls necessary for test control purposes.

(2) Courseware embedded tests/material (stored only on servers).

TRADOC Pam 350-70-5

- (a) Common controls.
 - (b) Secure electronic files through password protection and secure passwords.
 - (c) Restrict access when viewing material on-screen.
 - (d) Allow NO unauthorized copying/printing/ transfer/storage of files.
 - (e) If possible, randomly assign test versions and various versions of individual items/sets to learners.
- (3) Portable diskette/disk-based test material (floppy/CD/.zip files).
- (a) Common controls.
 - (b) Do not store or transfer material to an uncontrolled system.
 - (c) Password protect all files, or physically secure disk/diskette under single lock to electronically secure storage medium (diskette/Compact Disk-Read Only Memory).
 - (d) Restrict access when viewing material on-screen.
 - (e) Allow NO unauthorized copying/printing/transfer/storage of files.
- (4) Paper and pencil tests.
- (a) Common controls.
 - (b) Keep all copies of sensitive materials under lock.
 - (c) Make sufficient copies immediately before first administration. Make additional copies, only if necessary, and immediately before subsequent administrations.
 - (d) Make minimum copies required for single administration.
 - (e) Extraneous/unneeded materials are shredded or burned to destroy.
 - (f) Randomly assign alternate forms.
 - (g) Allow NO unauthorized copying/scanning of material.

Actions for loss, compromise, or possible compromise of sensitive test materials

j. The following guidance is provided if test compromise is suspected:

(1) Investigate every incidence of suspected unauthorized disclosure of sensitive test material, and substantiate, refute, or leave unsubstantiated the compromise. If the possible compromise is refuted, no further action is necessary.

(2) As the department/division head/commander, or other designated authority determines necessary, report the compromise/potential compromise to your chain of command. If needed for assistance, also report the compromise/potential compromise to the proponent school. The test proponent may advise on appropriate procedures to mitigate the risk.

(3) The commander/designated authority ensures that a thorough investigation of the compromise, possible compromise, or loss is made, and that proper actions are initiated to prevent recurrence of loss, or compromise of test materials. Additionally, appropriate authority should:

(a) Decide the risk mitigation factors to employ.

(b) Maintain a record of the results of the investigation and actions taken, if any.

(c) If warranted, initiate investigation under the provisions of [AR 15-6](#).

(4) If the compromise is substantiated, or not definitely refuted (that is, suspected but unsubstantiated), a risk assessment is immediately performed (based upon the level of control required of the test), and any serious consequences from the loss mitigated. At the discretion of the department/division head, or other designated authority, the procedures for mitigation should include, but are limited to, one or more of the following:

(a) Withdrawal of the test from use.

(b) Retesting of one or more learners, using uncompromised/unsuspected versions.

(c) Requesting assistance from the proponent school.

(d) Taking no action, if compromise is unsubstantiated.

Actions for destruction, transfer, and loan of sensitive test materials

k. Use the following guidelines for the destruction, transfer, and loan of sensitive test materials.

(1) Destruction of test materials. The test administering activity should immediately destroy test materials no longer needed, IAW such procedures deemed necessary, based upon the medium of the material (that is, burning, purging of files, reformatting diskettes, etc.).

(2) Transfer of test materials. Test administering activities may transfer surplus tests to another activity that needs them. The activity commander/other designated representative approves and monitors all transfers of sensitive test materials.

(3) Loan of test material. When a need arises, the activities may borrow a common use test from the nearest active duty, National Guard, or Army Reserve activity. In such cases, the activities involved take proper security precautions in transferal of the test material.

Actions for test control during administration

l. In addition to the guidance in paragraph k, above, take the following actions during test administration:

(1) Follow the proponent administrative instructions for each test precisely during actual administration, as the primary means of test control.

(2) Specific directions for test administration and scoring are contained in manuals that accompany each test, or in the learner assessment plan or test administration plan. Strictly follow these procedures. Test examiners and proctors should only use proponent test materials in preparation for, or during, the administration of test.

(3) If (in the opinion of the designated authority) any aspect of test security or administration is unclear or unmanageable, test administrators should immediately contact the test proponent for additional guidance, before administration.

8-3. Conducting test reviews and providing test feedback.

Purpose of test reviews

a. The purpose of a test review is to improve learner performance. In its most effective form, it provides constructive advice, direction, and guidance to learners in their efforts to raise performance levels. The test review is also used as a device to reinforce learning. As the instructor, take every opportunity to use the review as a means of clarifying, emphasizing, or reinforcing instruction in certain areas.

b. The intent of test reviews is to serve two purposes:

(1) Inform. Informational reviews serve to correct learner errors. Informational reviews should always be motivating, but motivational reviews do not necessarily provide information. A pat on the back or a word of encouragement may motivate a learner, but provide no information about the errors in performance.

(2) Motivate. Motivational reviews encourage learners to try harder.

c. It is important to realize that a positive critique is almost always seen as warmer and more sincere than a negative critique. Although you want to stress positive aspects of learner performance, it is also necessary to communicate weaknesses and needed improvements. An emphasis on the positive should enhance learner acceptance, and generate attempts to apply the information. The learner is less defensive and more willing to accept criticism. See [JA 350-70-5.8b](#) for additional information for conducting a test critique.

Characteristics of an effective test review

d. An effective test review has the following characteristics:

(1) Acceptability. Establish rapport and mutual respect. Learners that accept you are more willingly to accept criticism. They are confident in your qualifications, teaching ability, sincerity, competence, and authority. Your manner, attitude, and knowledge of the subject, along with conviction and sincerity, are accepted far more than your rank or position.

(2) Objectivity. An effective test review focuses on the learner's performance; not on personal opinions, likes, dislikes, and biases. The learner's personality and opinion are not at stake; nor should the learner have to agree or disagree with your beliefs. Although there is a need for openness and honesty, you do not have license to ridicule or show anger.

(3) Constructiveness. Constructive criticism points toward improvement, or a higher level of performance. Accompany all identified faults or weaknesses with positive guidance for improvement.

(4) Flexibility. Use flexibility when critiquing a learner's capability. The problem of selectivity is often confronted (that is, what to say, what to omit, what to stress, and what to minimize). Vary the organization and method of critiquing, according to the situation. Adapt tone, technique,

method, organization, and content of the test review to the occasion, and to the learner. To determine the best approach, consider the class situation, learner ability, subject matter, and the time allotted for the test review.

(5) Organization. Unless the test review follows some pattern of organization, valid comments may lose their impact. Almost any pattern is acceptable, if both you, and the learners, find it logical and easy to understand. As an example, start with general comments, continue with a group critique, and finally, request individuals to comment on their performances. Whatever the organization, prepare to change it, if the learners cannot follow the test review.

(6) Comprehensiveness. Make the test review comprehensive; the length depends on whether a discussion of major or minor points is desired. For effectiveness, feedback should include both strengths and weaknesses; with a determination of the best balance between the two. Comments confined to the excellence of the learner's performance are a disservice; also discuss those areas that need improving. Be specific with comments and recommendations. At the end of the critique session, learners should have no doubts concerning what they did well, what they did poorly and, most importantly, how they can improve.

Definitions e. See the glossary for definitions of [test reviews](#), [remediation](#), and [retests](#).

Conducting the test review f. The conduct of a test review, after each test, is a mandatory element of instruction.

(1) Conduct a test review on individual test items missed, even if all learners mastered all objectives.

(2) Give feedback to each learner on every item they missed (ensure you include every missed item, and provide feedback, either individually, or in a group setting). The recommended procedure is to provide a form to each learner, indicating only the questions that individual learner missed; then review each question missed by any learner. This should include "working" through problems.

(3) Recommend using methods to cover the items missed, other than returning of the actual test/test "booklet."

(4) Since the test review is a critical learning activity, give time and care to its preparation.

(5) During the test review, it is also highly recommended to capture any learner(s) comments that might indicate the need for test/instructional improvement. Involvement of a training/test developer is recommended.

(6) Slight modification to the above procedures is necessary when the instruction and test review is conducted via self-paced instruction. For example, for computer-delivered self-paced instruction, program test review and remediation into the course design and testing strategy. Allow an unlimited number of retest/remediation cycles, until the maximum time to complete the lesson is reached.

Remediation g. Remediation on every TLO not mastered on the first test administration is mandatory.

(1) The activities involved in remediation should directly address only the TLOs not mastered.

(2) The number of TLOs to remediate is the primary determining factor on the extent of remediation before retesting.

(3) In general, remediate the learners until they are confident of mastery of the objective.

Retests h. After remediation, at least one retest on each of the TLOs not mastered is mandatory. Make the decision to allow subsequent remediation and retests after the initial retest, once consideration is made of the following:

(1) The impact if further retesting is not allowed (that is, recycle, course expulsion, reassignment, reclassification, nonpromotion, Army expulsion, etc.).

(2) The resources expended thus far, to get the learner to this stage of their training.

(3) Availability of sufficient resources (including time) to conduct further remediation and retesting.

(4) Any evidence to suggest that the learner(s) will or will not master the objective with further remediation.

(5) Other exigent circumstances that preclude further remediation and retesting.

(6) It is highly recommended to make human decisionmakers, most knowledgeable of the above factors, the final deciding authority on decisions to stop/continue retesting.

(7) Document in the SEP the exact number and timing of retests, and how decisions to stop/continue retesting were made. Provide to each learner before administering the first test.

8-4. Quality control criteria for test administration. Quality control criteria include:

- a. The test administering activity reviews of the test guidance, and controls the test proponent provided for completeness and clarity.
- b. Strict adherence to the controls and procedures the proponent provided, and those contained in this pamphlet.
- c. Ensuring test results and comments get returned to the test proponent.
- d. Conducting test grading and reviews promptly.
- e. Scheduling test reviews and remediation sessions (as necessary) to allow adequate time.

**Appendix A
References**

**Section I
Required Publications**

AR 15-6
Procedure for Investigating Officers and Boards of Officers

AR 350-1
Army Training and Education

AR 380-5
Department of the Army Information Security Program

TRADOC Pam 350-70-4
Systems Approach to Training: Evaluation

TRADOC Pam 350-70-7
Systems Approach to Training: Courseware Design and Development

TRADOC Pam 350-70-10
Systems Approach to Training: Courseware Validation

TRADOC Reg 350-70
Systems Approach to Training Management: Processes, and Products

Section II

Related Publications

Air Force Handbook 36-2235, Volume 12
Information for Designers of Instructional Systems Test and Measurement Handbook
(<http://www.e-publishing.af.mil/pubfiles/af/36/afh36-2235v12/afh36-2235v12.pdf>)

MIL-HDBK 29612-1A
Guidance for Acquisition of Training Data Products and Services
(<http://assist.daps.dla.mil/docimages/0002/98/46/296121.PD0>)

MIL-HDBK 29612-2A
Instructional Systems Development/Systems Approach to Training and Education
(<http://assist.daps.dla.mil/docimages/0002/97/82/296122A.PD1>)

MIL-PRF-29612B
Training Data Products Performance Specification
(<http://www.dtswg.org/documents.htm>)

NOTE: Call DSN 680-5574 for any problems accessing a site.

Appendix B

Setting Test Standards (Passing Scores) for Tests

This appendix provides guidance for the setting of passing scores (NO/NO GO criteria) for knowledge-based tests. Use this guidance when using several test items to measure the objective evaluated, and the standard in the objective is not specific as to what level of mastery is required.

B-1. Standards for criterion-referenced testing.

a. For CRT, the passing (“cutoff”) score is set by objective, based upon the criterion standard for each objective tested within the test instrument. In other words, the item or set of items, measuring a single objective, is a separate test, with its own “passing” score. For ease of administration, a single test instrument is made up of many “mini-tests” that are administered concurrently within that instrument. (As an analogy, each

TRADOC Pam 350-70-5

“mini-test” is similar to a hands-on (performance) “station” used for certain performance tests, with each “station” a “mini-test,” measuring a different hands-on task/skill/performance.)

b. For example, see the test plan sample in TRADOC Reg 350-70, [section VI-7-6](#), paragraph e. In this example, it is determined that 10 questions are needed (in each version of the test) to adequately cover ELO 0001A. This set of 10 questions—and only this set—provides the information needed to make judgments about each learner’s mastery of the ELO. Therefore, the “standard” for this ELO is based upon the learners’ responses to this set of questions, and only this set of questions.

c. Applying the guidance of TRADOC Reg 350-70, [section VI-7-6](#), paragraph f, the “cutoff” score is reasonably set at 10 correct of 10 (100 percent), as long as the consequences of not passing the first time are reasonable (that is, the system can tolerate the time and resources needed for remediation and retesting). **NOTE:** This is especially true of self-paced instruction. In most cases, the objective criterion standard (passing score) for training delivered via self-paced instruction is set at 100 percent per objective (since time and resources are usually not a major factor).

B-2. Test construction errors. As initially written, some written test items, or set of items, may contain test/test item construction errors that may lead to a learner’s incorrect response, when the learner is actually a TLO/ELO master/performer. Proper test and instructional material validation should reduce these occasions. If test construction errors are noted after administration, zero-weight the item(s), to allow attainment of the assigned cutoff level, without remediation and retesting. The test proponent should give the administering activity specific, limited discretionary zero-weighting authority in the test administration instructions; or ensure the zero-weighting decision is coordinated with the test proponent after administration.

B-3. Standards for performance-based tests. Do not set the overall passing score at an arbitrary percentage (for example 70 percent) for a composite test, which tests many LOs, since this is extremely poor testing procedure. It is strongly recommended to set the standard for performance-based (written) tests 100 percent. Only set the standard at less than 100 percent if—

a. The 100 percent standard is not feasible, due to the resources expended to remediate and retest to a 100 percent standard, and there is clear and convincing evidence that less than a 100 percent standard will distinguish between masters and nonmasters, or is adequate to certify competency.

b. The clear lack of criticality of the specific TLO/ELO tested allows a less than 100 percent standard, and there is clear and convincing evidence (from a consensus of SMEs and test developers) that less than a 100 percent standard will distinguish between masters and nonmasters, or is adequate to certify competency.

c. The test measures a supporting skill/knowledge for a performance that is trained later in the course, for which less than perfect knowledge/skill does not significantly detract from the effectiveness of the later training (that is, the lack of perfect knowledge of the steps in a process are overcome by later training and practicing of the entire process).

Note: This philosophy is new to those previously taking or giving “norm-referenced” or “domain-referenced” tests throughout their years of schooling. These tests usually accept less than perfect knowledge/performance, because teaching to mastery cannot be accomplished or is even desired. Within Army training, we cannot afford not to teach to mastery. When performance/knowledge must be guaranteed due to the consequences of inadequate performance/knowledge, testing to the very highest possible standard - 100 percent when necessary - is absolutely critical. For example, which single critical knowledge (fact) regarding the characteristics of a particular bomb fuse would you like to not have to prove you knew if you later had to defuse a bomb? The same could be said of knowledge of the actual procedure that should be used for disarming a particular bomb before time/resources are wasted in practicing the disarming.

Appendix C

Rank Ordering Learners

C-1. Using available tools.

a. The purpose of this appendix is to provide ways to make norm-referenced decisions about learners, without developing and using NRTs. In particular, the decision frequently made is how a learner “ranks” in relationship to the other learners in the course/class. In order to make such distinctions, it is necessary to find data, which more precisely discriminates between learners, than the limited data available solely from a GO/NO GO decision.

b. It is Army policy not to construct learner performance measurements/tests for the sole purpose of rank ordering learners. First and foremost, use the scores collected for a learner from a measuring instrument, to determine if the learner met the mastery standard. The learners who meet or exceed the mastery (cutoff) score for the objective are given a GO for that TLO/test. Learners not meeting or exceeding the mastery standards are given a NO GO. There is no “almost met,” “grading on the curve,” “the top 10 percent of the scores are given an A,” or “everyone must make 70 percent on each test in the course,” etc. Possible ways to rank order students, without degradation of the intent of the assessment instrument, are discussed in the following paragraphs.

C-2. Methods of collecting rank ordering measures.

a. General criterion for rank ordering. Determine a rank ordering of learners from both objective data (obtained either directly or vicariously from test scores, times, or

TRADOC Pam 350-70-5

other objective data) and subjective data (opinions of instructors, other learners, or from each learner.

(1) Objective source—raw test scores. If the test scoring results in a range of scores of any nature, after the mastery standard is applied and the GO/NO GO decision is made, use these scores to discriminate among (that is, rank order) learners. For example, if the GO criterion is 4 of 5, then those scoring 4 of 5 could receive a score of 80 percent, and those scoring 5 of 5 assigned a score of 100 percent. **(Note:** Assigning a percentage of correct, versus total attempted, is an arbitrary numerical assignment. Use any assignment which yields relatively sequential numbers (that is, an assignment of 4 points to the 4 of 5, and 5 points to the 5 of 5). Likewise, those scoring 3 of 5 could receive a score of 60 percent. In this last case, the learner did not reach the mastery cutoff level (4 of 5) and will receive remediation. For rank ordering purposes, the score of 60 becomes the learner's final score for rank order "averaging" or assign the learner the score achieved on the retest. The above procedure works only if the learner is forced to make five attempts, and not stopped once the criterion is reached (performs correctly the first four times). Forcing a learner to continue testing after the mastery standard is reached, just to produce a relative percentage score, is wasteful, and not advised.

(2) Objective source—time to complete performance test. Another measure to use for rank ordering is used when the performance has a time standard (that is, correctly assemble/complete/disarm/emplace/request/identify/compute, etc., within X minutes). The time that each student takes is recorded, and the learner's "time to standard" is rank ordered. The mastery standard applied is an absolute (that is, within X minutes), but the relative times to mastery provides discriminating "ranking" data. However, a major flaw can occur with this procedure. If the learner knows that the quicker a performance is completed results in a higher ranking on that performance, a fully capable learner might unnecessarily rush through the test/item, and make mistakes they otherwise would not make. Therefore, a learner that demonstrates mastery, may require remediating/retesting, just because they wanted to rank higher than another learner on the instrument.

(3) Objective source—time to complete knowledge-based test. Although not usually part of the standard, if desired, collect the time to completion for knowledge-based tests. Learners are ranked higher the earlier they finish. The problem with this method, of obtaining objective relative rank ordering data, is that the learner finishing first is not necessarily the "better" learner. The "better" learner might carefully check their responses before turning in the test. If using this method, be aware of this problem. Also, the situation described in paragraph (2), above, of the learner rushing through and making critical mistakes, just to quickly finish, may also occur for the knowledge-based test. This situation does not give a true indication of the learner's ability, and as a result, if the criterion is not met, results again in remediation and retesting.

(4) Objective source—time to complete self-paced instruction. It is also possible to collect the absolute time (time-on-task) it takes a learner to complete a unit of self-paced instruction (all the instruction, not just the test). Compare this time with those from other learners, to rank order the learners. Although a “rank-order by total time to complete” is obtained, the implied interpretation of this acquired rank order—in particular, that learners who finish sooner are in some way “better” than others taking more time—is a flawed assumption. Learners learn at different rates/paces, and self-paced instruction uniquely provides the opportunity for the learner to learn at their own comfortable pace. Those finishing sooner are in no way “better” than those that finish later.

(5) Objective source—assignment of points to attempts. A possible way to produce an objective score, from a dichotomous GO/NO GO decision, is to assign points by number of tries to mastery. A raw score is not necessary just the number of tries to mastery; it is possible to assign “points” to all students who are first time GOs, and lesser/no points to those achieving a GO on later tries. One example is that all first time GOs receive 2 rank-order points, second time GOs receive 1 point, third time GOs receive 0 points, or some variation thereof. These “assigned scores” are summed or weighted, and summed to produce a rank ordered list. However, this method will not necessarily make the very fine distinctions required to identify a small number of “top” learners, since many learners are likely to achieve mastery the first time they try.

(6) Subjective source—observation ratings of attributes. Use the observation as a subjective measure for rank ordering learners. For each learner, the observer develops and evaluates a detailed list of one or more carefully selected, scalable attributes to observe. The most likely observers are the instructor(s), course/class manager, other learners (peer ratings), or the learner themselves (self-ratings). The key to implementing this method is having clear criteria to observe, and having the instrument scaled to make the fine rank-ordering discriminations needed. The learner attributes/traits to observe and “evaluate” may include such attributes as “leadership ability,” “military bearing,” “command presence,” “adaptability,” “attitude,” “timeliness,” “helpfulness,” “situational awareness,” “safety adherence,” “participation,” “cooperation,” “enthusiasm,” and “communication skills.” (**Note:** Although less objective than methods above, this method is likely the most viable, since it avoids all the difficulties inherent in the methods above.)

b. Combination of techniques/sources. Within a single module, phase, or course of instruction, it is possible to use several of the sources listed in paragraph a, above, to make the final rank order determination. For example, rank-ordering the student on each test from top to bottom, using any of the objective sources (although the “tries to mastery” technique usually yields no more than three levels), and rank-ordering using the trait assessment. The learner (and ties) at the top level of the ranking for each test is assigned the most (or the fewest) points, those at the next level fewer (more) points, etc., until all learners are assigned points. After summing these points, those learners with the greatest (or less) points (and any ties) are declared honor graduate(s)/commandant list. Use test weights in this process.

C-3. Rules for rank ordering. The following rules apply to making rank order decisions about learners.

a. Do not develop a test solely for the purpose of making fine distinctions among learners for rank ordering purposes.

b. Do not select a particular type of test, based upon its ease in making fine discrimination decisions for rank ordering learners.

c. Use data from performance tests for ranking purposes, only if each learner is assigned a meaningful “relative score” (as compared with other learners) from the test results.

d. If available, use data normally collected from performance tests, that when interpreted, could make fine discriminations useful for rank ordering.

e. Use time-to-completion data from performance or knowledge-based tests judiciously since it may not be a good measure of relative knowledge/ability, and may lead to initial unnecessary failure for learners who make errors due to unnecessarily rushing through the test.

f. If using observations to measure traits/attributes, ensure that:

(1) The instrument explains and describes fully the attribute measured.

(2) The instrument has a clear, anchored scale upon which the observers can “rank” a learner.

(3) The observer has the opportunity to observe the attribute/trait in each learner.

(4) The observer has the skill to rank an individual on the attribute/trait.

(5) Errors of bias in ratings are determined and controlled as much as possible.

C-4. Weighting of tests/test items for rank ordering. Any method above that produces a learner-discriminating numeric value (that is, a relative scale score)—whether objectively or subjectively derived—can use an “assigned test weight” to further discriminate between learners for rank order purposes. Subsequently multiply and sum these numeric relative scale scores/points by the “assigned weight” of the test/item/attribute, to produce an artificial “weighted score,” which is used to produce the desired rank order for learners on the module/phase/course. **WARNING:** This assignment of points and weights, the weight sums, and the subsequent rank order is not reliably interpreted that one learner is definitively “better” than another on a test, task, or observed attribute, or “achieved more” in a module/phase/course.

Appendix D

Automated Tools for Test Development

D-1. Using this appendix. This appendix provides various links to test authoring tools to assist in developing web-based assessment tools. As an Army requirement, most ICW tests are Sharable Content Object Reference Model (SCORM) compliant. For IMI courseware, ensure the test design tool can develop SCORM-compliant tests, and there are no licensing restrictions on the use of the product, or the tests created using the product. **Note:** Listing of products below does not constitute Army endorsement.

D-2. Authoring tools.

a. Hot Potatoes is a low-cost software package to use when creating multiple choice, fill-in-the-blanks, crossword puzzles, and more (<http://www.halfbakedsoftware.com>).

b. Quizmaster is a great free generator that offers simple exercises and also some millionaire-style games. To generate the exercise, complete the information on the website, and click a button (<http://cybil.tafe.tas.edu.au/~capsticm/quizman/qmhome.html>).

c. To provide assessments (multiple choice, true/false, etc.) over the web, with the ability to track scores, Course Builder for Dream Weaver, 12teach, and Authorware are some good programs to utilize. Additionally, look at EasyT Survey and Assessment at <http://www.globalpresence.com.au/easyt/products>; and access an online sample at <http://www.globalpresence.com.au/showcase>.

d. DazzlerMax (http://www.maxit.com/daz_templates.htm) is an award winning, electronic-learning authoring tool to use for rapidly creating interactive training, testing, and presentations—**FAST**—without having to learn a complex scripting language, or compromise instructional design sophistication. DazzlerMax capabilities include:

(1) Interaction where the user performs a drag-and-drop operation, and the operation is scored.

(2) Performance of a software simulation, complete with capturing mouse clicks and keyboard selections. The user enters data, validates the data, and is able to use that data later in their interaction.

e. Quiz Rocket is a unique, easy-to-use quiz and survey program. Unlike other web-based tools, Quiz Rocket creates an interactive, media-rich environment for web users. Using Quiz Rocket's fill-in-the-blanks, template approach, you can customize quizzes and surveys around any content, and publish them on the Web for access anywhere, anytime. Quiz Rocket's template design lets users design quizzes and surveys around any subject or topic. Six question formats are available: multiple choice, sequencing, matching, true/false and yes/no, short answer, and branching. Mix and match test and survey questions in the same quiz. Use data collection fields to gather demographic

information. One program creates an unlimited number of tests and surveys, each with up to 200 questions (<http://www.learningware.com/quizrocket>).

f. RapidBuilder v3.0 Deluxe is a special advanced edition of RapidBuilder v3.0, XStream's 3x award-winning, 100 percent Programming-free Simulation-Authoring Technology for creating, customizing, and deploying fully interactive, multimedia simulations that realistically duplicate the functionality of the live Windows environment. RapidBuilder v3.0 Deluxe includes all of the familiar RapidBuilder features and capabilities, including continuous capturing of live software applications, multimedia edit suite, featuring appended/inserted frames, text, graphics, audio files, recorded sound clips, embedded sound tracks, video, navigational hyperlinks, links to external files and web pages, menus, mouse pointer animations, and much more (<http://www.xstreamsoftware.com>).

g. AIMS-PC is an evolving system developed to facilitate training management functions at the schoolhouse, through automation of such tasks as inprocessing, student management, graduation, and outprocessing; maintaining test item banks and test versions; grading tests and maintaining gradebooks; and reports generation, to include diplomas and course certificates. Future capabilities will include integration of the Program of Instruction Management Module functions to support student management in the training units, and a scheduling module to assist in scheduling individual training (<http://www.aims-r.army.mil/>).

h. Saba Publisher allows course developers to select from different question types, and then quickly and easily assemble a test. Question types Saba Publisher supports include essay, short answer, fill-in-the-blank, multiple choice, true/false, drag-and-drop, and matching. In addition, course developers can choose to randomize the questions within a test, set a specific amount of time for the user to complete the test, and define the passing grade for the test. The Saba Publisher test wizard also allows the flexibility of creating either tests, based on the cumulative knowledge gained from a course, or section tests the user must pass to continue on to the next section within a course. Tests are compatible with the Army's implementation of the Saba Learning Management System (http://www.saba.com/english/products/pdf/Saba_Publisher_v5.pdf).

i. AIMS-PC uses a commercial off-the-shelf product called LXR Test (<http://www.lxrtest.com>). LXR Test provides an objective/test item bank, which allows development and administering of performance based tests. LXR Test also prints and presents exams, captures electronic and mark-sense responses, scores exams, posts grade books, and facilitates the evaluation of training effectiveness. The AIMS-PC contains an import/export feature that allows users to export class rosters from the AIMS-PC into LXR Test, and then import grades from LXR Test back into the AIMS-PC.

Appendix E Review and Revise Learning Objectives

E-1. Analyze learning objectives.

a. For suitable use in developing test items, an objective should contain explicit statements of performance, conditions, and standards. Since objectives are not always “neatly packaged,” some searching is often necessary to find the performance required, organize applicable conditions and standards, and express them in terms of performances to observe.

b. Analyzing objectives requires examining each part separately. Paragraphs (1) through (3), below, divide the sample objective in figure E-1 into the critical task required to perform the job, the expected conditions of performance, and acceptable standard for performance.

Using an M543 wrecker and an M-16 sling, the wrecker operator trainee will be able to operate the hoist, as directed, in unpackaging the Honest John Warhead section, following the sequence specified in TM 9-1340-202-12. Performance will occur on an outdoor, flat, hard surface.

Figure E-1. Sample learning objective

(1) **Task.** Operating the hoist. The main intent of the objective (operate the hoist) is directly observed, and needs no indicator.

(2) **Conditions.** Several conditions are stated throughout this objective, rather than clustered in one part. First, it specifies the equipment to use. Second, the material operated on (the warhead) is specified. Third, the environmental conditions are described. And finally, special instructions are implied (the trainee is directed in the operation of the hoist). This objective uses all four types of condition statements (that is, what is available to work with, what is worked on, environmental circumstances, and limitations/special instructions).

(3) **Standards.** In the sample objective, the standard is the SOP type. In order to satisfy the objective, the trainee follows the sequence specified in the appropriate technical manual for the Honest John Rocket System. The learner completes all steps in sequence. No time standard is suggested in the objective, but it is inferred the trainee performs the task within reasonable time limits.

c. An objective is not automatically suitable for test development purposes because it contains the essential three parts. To ensure the adequacy of the LO, follow the steps outlined in the following paragraphs.

E-2. Assess the adequacy of the task (performance). Use the four major checks in paragraphs a through d, below, to assess the adequacy of objectives. Work from the

TRADOC Pam 350-70-5

list of objectives, broken down into their three parts (task, conditions, and standards), to facilitate these checks.

- a. Ensure each task covers a single task, and is not a combination of tasks (unitary).

(1) It is important to use unitary objectives to develop a test (each covers one task only). It is more difficult to write test items for compound objectives (those covering more than one task). Before proceeding, break down any compound objectives into unitary objectives.

(2) To check that objectives are unitary, examine the task statement. Look at the performances the objective calls for, and ask the following questions:

(a) Does each task call for performance on just one task?

(b) Are all tasks independent? (That is, successful performance on one objective does not require successful performance on a preceding objective.)

If the answer to either question above is a definite “no,” the tasks are probably not unitary. Carefully subdivide the tasks into unitary tasks, as appropriate. Submit the list of unitary objectives, through channels, to their originator for verification.

(3) When subdividing compound objectives into unitary objectives, break down the “task” (performance) part of the compound objective. Each unitary objective may include the same conditions and standards specified in the compound objective from which it was derived. Change conditions and standards when the conditions and standards for the unitary objective differ from the compound objective.

- (4) Revise tasks that are not unitary.

(a) The sample tasks in figure E-2 are not written as unitary tasks. The sample objectives 5 and 6 call for performance on several different tasks, while the other objectives concern single tasks. In addition, there is a lot of overlap (that is, lack of independence) among objectives. For example, objective 5 requires controlling arterial bleeding, while treating for shock is probably common to all objectives.

- | |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ol style="list-style-type: none">1. Treat for shock.....2. Treat for nerve gas inhalation....3. Administer mouth to mouth resuscitation.....4. Control arterial bleeding.....5. Give first aid for burns, chest, abdominal, head, face, and neck wounds; and open arm and open leg fractures...6. Correctly apply a tourniquet and construct a hasty litter. |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Figure E-2. Sample of tasks that are not unitary

(b) The steps in figure E-3, below, represent one example of revising the objectives in figure E-2, so that each task is a single task, rather than a combination of tasks. The rewritten objectives are now nearly independent, and cover separate, single tasks (unitary). Note that applying a tourniquet is incorporated in objective 6 (see fig E-3). It is not really a separate task, but a normal part of treating compound fractures, where blood flow is not otherwise controlled. Although objectives 5 and 6 may each seem to cover several tasks, they actually do not. First aid of head, face, and neck wounds is one task—the procedures do not differ. The procedures for treating open arm and open leg fractures are also the same. All tasks in the original six objectives are now shown in a unitary fashion in the eight new objectives. No performances were changed—only divided into unitary performances. The conditions and standards for each objective may require changing, because of the redefinition of the tasks.

- | |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ol style="list-style-type: none"> 1. Treat for nerve gas inhalation... 2. Give first aid for burns... 3. Give first aid for chest wounds... 4. Give first aid for abdominal wounds... 5. Give first aid for head, face, and neck wounds... 6. Treat open arm and open leg fractures (bleeding cannot be controlled by direct pressure, digital pressure to pressure points, or elevation). 7. Construct a hasty litter. 8. Administer mouth-to-mouth resuscitation. |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Figure E-3. Sample of unitary tasks

b. Ensure the main intent of each task is clear (clarity).

(1) Every task statement contains a main intent, and an indicator. The main intent is the statement of the task that tells what the objective is mainly about (that is, the skill or knowledge the learner is to develop, or the performance, which is the purpose of the objective. In most instances, the main intent and the indicator are the same). A main intent is either overt (observable), for example, “disassemble an M-16,” or covert (unobservable), for example, “recognize the differences in appearance between poisonous and nonpoisonous snakes.” If covert, add an indicator to the objective to tell how to evaluate the main intent.

(2) The indicator is the action verb of the objective’s task statement, when performance of the main intent is not inferred, and not in itself directly observable. For example, if the main intent is “Discriminate the shears used for cutting a straight line, and those used for cutting a curved line,” include an indicator such as “by circling the picture of shears used for cutting a curved line.” In this case, the main intent (“discriminate”) is covert, or not directly observable, requiring the addition of an indicator.

(3) Look at the performance statement for the objective, to ensure that the main intent of the objective is clear. If the performance statement calls for the performance that demonstrates the skill, the main intent of the objective is clear. If the performance

statement does not include the performance that demonstrates the skill, perhaps the performance called for misses the main intent of the objective, or possibly does not provide directly observable performance. In either case, ensure that the main intent is clear, or defined operationally.

(4) A performance statement is usually unclear if the intent/indicator is not observable, or measurable. Nonspecific phrases such as “be aware of,” “demonstrate an understanding of,” and “know” require close examination. In the phrase “demonstrate an understanding of the differences between treating a simple fracture and a compound fracture,” the main intent is unclear, and the purpose of the objective is not known. It could intend to find out if an individual can treat both types of fractures; or distinguish if a compound fracture is treated like a simple fracture. Also, there is no indicator to figure out how to measure the “demonstration of an understanding.” At first glance, the statement appears to state a performance, but the learner is left guessing what performance is required. Ensure each task has a main intent and an indicator (that is, Main Intent = Overt/Covert + (Indicator)), as shown in figure E-4.

Verb	=	Overt/Covert	+	Indicator
List	=	Overt	+	List
Analyze	=	Covert	+	In writing
Recognize	=	Overt	+	Point out, select, state, or identify

Figure E-4. Task components

(5) The following example of a performance statement has a clear indicator, but the main intent is unclear: “Demonstrate an understanding of good briefing skills, by listing the three main parts of a briefing.” The indicator is clear and calls for an observable act—listing. The main intent may seem clear, but “listing the three main parts of a briefing” does not demonstrate an understanding of good briefing skills. It only indicates knowledge of the three parts, not the ability to conduct a successful briefing, or even to recognize whether a particular briefing is organized in three parts. Although the main intent is stated, it is not clear. The learner does not know the main intent, and the indicator does not provide help in interpreting it. The indicator may reflect the performance the writer of the objective wants measured, and the main intent is poorly stated. Or, if the indicator is poor, use of a different indicator will clarify and support the main intent.

(6) Revise tasks with unclear main intents.

(a) Clarifying the main intent is the most difficult part of this process. Unclear main intents are most prevalent with “soft skill” areas, such as leadership or tactics. Yet these areas, because of their importance, require the student to demonstrate these skills. The designer should operationally define the main intent. For example, the task: “Supervise the maintenance of a unit” does not have a clear main intent. The word “supervise” may have different meanings to different individuals. However, if “supervise” is defined as “the planning, directing, monitoring, evaluating” of an activity, then the meaning of the word “supervise” is the same to each student.

(b) There are other unclear main intents that are not operationally defined, such as “know,” “understand,” or “be aware of,” since these are not action verbs. In such cases, go back to the analyst, job incumbent, or supervisor, and determine the original meaning of these terms. The statement itself may contain a clue. For example, “know how to conduct an active defense” contains an indication of the intent (that is, “conduct an active defense”).

c. Ensure performance indicators are simple, direct, and part of what the trainees can already perform.

(1) If the main intent of the objective is clear, next determine if it is overt or covert.

(a) An overt main intent is one which is observable and measurable. Overt main intents do not require indicators, they already tell what performance is called for, and how to measure it.

(b) Covert main intents require indicators, since the required performance is not directly observable. A covert main intent is an action verb, but the performance is unobservable, and the indicator tells how to measure whether or not an individual can perform it.

(2) If an objective’s main intent is measured through an indicator, ensure that the indicator is appropriate. A good indicator is:

(a) Simple. Keep it as uncomplicated as possible. Do not obscure the main intent with an unnecessarily complicated indicator.

(b) Direct. Use indicators when the main intent of the performance statement calls for a performance that is either not directly observable, or not practical in the testing situation. The indicator should allow determination of whether or not the main intent was satisfied, without going through chains of inference.

(c) Part of the trainees’ normal repertoire of behavior. Ensure that the trainee is able to perform the indicator behavior; since the indicator behavior itself is not what is trained or tested. It is only used as a measure of the main intent. If the indicator is not a part of the trainee’s normal repertoire, two things are measured—performance on the indicator, and performance on the main intent.

(3) In the example of a performance statement “recognize the major bones of the human skeletal system, by drawing a picture of each bone beside the names of the bones,” recognizing bones is the main intent, while drawing pictures of bones is the indication of recognition. Drawing pictures of bones is a direct indicator in this case, since if a person can draw the correct picture next to the name of the bone, it is known the learner recognizes the bone, and no inference is made. But drawing a picture is not a simple indicator. Worst yet, drawing a bone, well enough for an examiner to identify it,

is not a part of the trainees' normal repertoire, unless the trainee is a skilled illustrator. A trainee might fail to satisfy the objective because of poor drawing, not because they did not recognize the bone. The indicator is poor for another reason—the main intent is to recognize bones, but the indicator requires the person to recall what the bone looks like, then draw it. A better indicator for this main intent is "...by writing the name of the bone next to the picture of the bone," or, even better: "...by choosing the correct name from the list provided, and writing it next to the picture of the bone."

d. Ensure each task contains an action verb.

(1) Check to ensure the use of precise operational terms to write the statements of performance, conditions, and standards, so that each statement is easily translatable into actions. For the statement of performance, check for clarity of the main intent, and appropriateness of the indicator. A further check on the performance statement of the objective is helpful at this point.

(2) Ensure every objective states precisely what the individual is to perform. Make the statement of performance clear enough for training and testing that performance. The following objectives include examples of stated performances:

- (a) Climb the telephone pole.
- (b) Disassemble an M-16 rifle.
- (c) State the conditions for which a tourniquet is applied.
- (d) Camouflage the helmet.
- (e) Add two 5-digit numbers.

(3) Every statement of performance in paragraph (2), above, includes an action verb. This verb is usually the key to the performance, and tells what to do. For example, in the statement in paragraph (2)(c), above, the action verb is "state." You can test the student's ability to state these conditions. If that statement of performance read "Appreciate the conditions for which a tourniquet is applied," knowing when a student "appreciates" the conditions is hard to test. Sometimes the action verb is not the key to the performance to train and test; it may only indicate the performance. When you cannot point to the performance itself, the action verb should specify the appropriate indicator of that performance. For example, consider statement (e), above. It is clear the performance is to "add." To know when someone successfully adds two numbers requires an indicator, since the act of adding is not observed. In this example, add an indicator to the statement of performance (that is, "Add two 5-digit numbers, in writing."). Although "writing" is the observable action, the main intent of the performance is adding, not writing. If the statement of performance calls for an action (includes a main intent) that is not directly observable, add an appropriate indicator.

(4) Ensure each statement of performance uses a specific action verb. Table E-1 provides examples of verbs often found in the performance statements of objectives. See TRADOC Reg 350-70, [appendix D](#), for a full list of verbs.

Table E-1
Examples of verbs

Non-Action Verbs	Specific Action Verbs
Appreciate	Brake
Be aware of	Check off
Be familiar with	Label
Know	State
Understand	Turn

Note: Even though the examples of verbs above are often used to specify performance in objectives, only those in the right column are suitable.

(5) Revise tasks that do not contain action verbs. After determining that a task does not contain an action verb, rewrite the task. Analyze the task statement to determine the intent. For example, the tasks “know the differences between...” “know the characteristics of...” or “understand the function of...” each contain verbs that are not action verbs, but common sense indicates the intent. Rewrite the tasks as: “state the differences between...” “list the characteristics of...” or “explain the function of...” The task analysis data might also provide some clue to the action verb. Study the elements of performance to determine the action verb. Return to the job analyst, incumbent, or supervisor, if required, to ensure each task statement contains an action verb and an object of that action.

E-3. Assess the adequacy of conditions.

a. Ensure statements of conditions and standards are complete, and written in precise, operational terms. Writing test items is dependent upon a description of complete conditions (see fig E-5). For example, to write a test item for the task “Run a mile,” relate the task to a job (which in this case, requires the job holder to run with a 40 pound pack, up a hill with a 10 degree slope, in less than 8 minutes). At this point in the analysis of the objective, there are two primary concerns: the conditions specified are complete; and the conditions match the task.

TASK + CONDITIONS = TEST ITEM(S)

Figure E-5. Equation

(1) Check the statement of conditions for completeness. The conditions of the objective should provide all the information required, in order to set up the conditions of the test. The conditions specify tools and equipment, special JAs, manuals,

TRADOC Pam 350-70-5

supervision/assistance (if any) received, special physical demands, environmental conditions, and location of performance of the task (see fig E-6). Ensure all this information is available, to properly construct a test. See table E-2 for examples of condition statements.

<u>Task</u>	<u>Conditions</u>
	Tools/equipment
	Manuals/references
	Supervision/assist
	Special physical demands
	Environmental
	-Light
	-Visibility
	-Precipitation
	-Temperature
	-Noise
	Location of performance

Figure E-6. Sample of complete list of conditions

**Table E-2
Examples of condition statements**

Task	Appropriate Conditions	Improper Conditions*
Repair 45 kilowatt (kW) generator.	Given a 45 kW generator with a broken shaft bearing.	Given a malfunctioning generator.
Replace a front handgrip.	Under ordinary field conditions, in daylight.	Under ordinary conditions.
Determine the flow of current.	Using a multimeter.	Using appropriate test equipment.
Assemble a model airplane.	Using M2A1 epoxy.	Using the right glue.

*In each of the improper conditions above, some additional information is needed, such as, the type of generator, the type of conditions, the right test equipment, and the right glue. If additional instruction is required, the conditions are incomplete.

(2) Check that statements of conditions match the task.

(a) Ensure the specified conditions match the task. In table E-2, above, the correct condition for the first example specified a 45 kW generator, because the task states: "Repair 45 kilowatt (kW) generator." If the task stated: "Repair generators," no requirement exists to specify "45 kW" in the conditions. The task is a different task, and therefore, an appropriate condition might state "given a malfunctioning generator."

(b) Ensure the condition statement specifies all conditions that may affect the performance of a task. For example, if the requirement is to repair the 45 kW generator "in total darkness," the performance of the task is affected. Another example is the construction of a foxhole. Europe and the continental United States uses entrenching tools to dig foxholes; Alaska uses shaped charges. While it appears the task is the

same, they are quite different. To construct test items that measure the performance of a task requires specifying all the conditions.

b. Revise statements of conditions that are incomplete, or do not match the task. Use the following sources of information to obtain the conditions:

(1) Examine the job analysis data. The job analysis data and job performance measures may provide clues to all of the conditions required to perform a task.

(2) Check applicable references.

(3) Contact the SME. Also, check with the SME when unable to spot defects in the condition statement, instead of guessing, or writing something in.

E-4. Assess the adequacy of standards.

a. Assess the standard.

(1) Each test item should specify the standard (criterion) by which performance is evaluated. In other words, ensure every standard indicates how well, or how quickly (or both) to complete a performance. Just as with tasks and conditions, clearly state standards in the objective, in order to know how to train or test. For example, in the objective: "Be able to type reasonably accurate, using a computer, under standard office conditions," the objective is lacking a clear statement of standards. With no standards for speed and accuracy, there is no indication of how fast to train learners to type, or how fast a learner would have to type to pass a CRT, in order to satisfy the objective. A complete objective might read: "Using a computer, in standard office conditions, type 50 words per minute, corrected for accuracy (one word per minute subtracted for each mistake)." This objective provides the standards to achieve in training, and the level of performance learners should demonstrate on a test.

(2) The following six specific types of standards indicate how well (quality) or how quickly (time) to perform an action, or complete a product. A complete test item should specify at least one of the six types of standards. Often, an item combines several types of standards; for example, it could include both quality and time specifications.

(a) Standard operating procedure (quality)—performance matches a specified SOP. This standard specifies that a performance is complete, and all parts of a performance performed sequentially. (Example: "Given a map with forward observers and enemy troop positions marked, the trainee must issue a "call-for-fire," using the sequence as specified in the U.S. Army Infantry School Operations Handbook.")

(b) Zero error (quality)—performance is completed to 100 percent accuracy (or product is completed exactly right). (Example: "The trainee will set the quadrant on a 4.2 mortar sight to a specified mil. He must set it at the exact mil (for example, 345) he is told." If the trainee is off by one mil, the standard is not met.)

TRADOC Pam 350-70-5

(c) Minimum acceptable level (quality)—performance meets a specified minimum acceptable level (or product meets specified tolerance). (Example: “Using a standard oral thermometer, take a patient’s temperature and record it, to the nearest two-tenths of a degree.” The minimum acceptable standard is the nearest two-tenths of a degree, not the nearest tenth.)

(d) Subjective quality (quality)—performance achieves certain characteristics, measured qualitatively (or product has certain subjective characteristics, for example, boots must have a bright shine). (Example: “Be able to land a UH-1D helicopter, with power off, using autorotation, and make a soft landing from 1,000 feet.” The standard of a “soft landing” is qualitative. Use care to define standards of subjective quality as precisely as possible, so that two observers would agree, in most cases.)

(e) Time requirements (time)—performance is accomplished at a certain minimum speed. (Example: “Correctly multiply pairs of 5-digit numbers, using a desk-top calculator.” The trainee is required to get the correct answer for at least 10 such multiplications per minute. It is important for the trainee to multiply quickly using this calculator, hence the time requirement. Words-per-minute is a similar requirement for typists.)

(f) Production rate (time)—performance yields a certain daily or monthly output (complete products at a certain rate). (Example: “A three-man wire team should be able to lay and splice in 3 miles of wire per day, over moderately difficult terrain, connecting at least three different locations.” In this example, the important achievement is the amount of wire laid per day, not the speed with which it is laid.)

b. Check to insure standards are complete.

(1) A complete standard specifies the precise nature of the output; the number of features the output contains; the number of steps, points, pieces, etc., to cover; and any quantitative statement that indicates the acceptable portion of the total.

(2) The output of the performance of a task is usually identified as a product. The actual process of performing a task is sometimes the product; or a combination of a process and a product. For example, the product of the task “List the principles of war” is a written list; the process of “playing a violin” results in playing each note; the process and product of “bake a loaf of bread” is the process of mixing and baking the bread, and the resulting product tasting like bread at completion.

(3) The number of features the output contains is a description of what the product looks like when completed properly; such as, the “list” should contain “principles of war.” If an output involves a process, it should include the number of steps to complete; such as, “the 9 steps in preparing the bread.” And lastly, a quantitative statement is required, such as “carry out all 9 steps, in sequence.”

c. Check to ensure accuracy of standards. Accurate standards indicate how close to correct (proper) performance is required, and the exact tolerances, values, or dimensions that an acceptable answer/performance assumes (these may be qualitative). Establish standards of objectives based on the job; other times the author of the objective arbitrarily establishes the standard. In either case, the standard is an absolute standard, where no error in performance is tolerated, or one in which some margin for error is acceptable. For example, an absolute standard for the task “run a mile” is in 8 minutes. This means, to successfully complete the task, a mile is run in exactly 8 minutes—no longer or shorter. Rather than using this type of standard, it is normal to allow a range, such as “from 7 to 9 minutes.” Provide tolerances (such as “ $3 \pm .2$ ”), values (such as, “the steak should be light to medium pink in the middle”), or dimensions (such as “the steak may be either ___ or _____”), with an objective device, which indicates if the trainee did, or did not, perform properly.

d. Revise the standard.

(1) Base revisions on what is wrong with the standard. Incomplete standards require looking at the task and conditions, and determining the outcome (result) of performance. For example, the outcome of the task “fire the M-16 A1 rifle” is a round of ammunition, going down range, and hitting a target. Ensure the standard states the target, and how many times to hit the target.

(2) If the standard is not accurate, add necessary specifics, such as, how close to correct performance, or the tolerance values and dimensions the product assumes. Obtain clues from the task analysis; references (such as field manuals, technical manuals, etc.), or from the SME.

Appendix F

Interactive Courseware Test and Measurement

F-1. Computer managed instruction (CMI). An important aspect of ICW development is test and test item design, and the design of CMI functions and records. Prior to designing ICW, review the selected authoring software, to determine the extent of data collection and analysis that is possible. Develop ICW tests to measure the intellectual skills related to, and associated with, each hands-on task or TLO. Computer managed instruction is the function of the ICW authoring software related to student test and measurement data collection. Table F-1 describes the functions included in CMI.

Table F-1
CMI administrative and performance tracking functions

No.	Administrative	Performance Tracking
1.	Registration of the student in an ICW course.	Employment of different types of test items (e.g., digitized video, and graphic and animated images).
2.	Point-of-entry for the student into the course, often based on a pretest performance, or previously “bookmarked” location.	Collection of data regarding the student’s performance metrics for test items and test segments.
3.	Students can leave a lesson and return to the same point at a later time.	Provision of immediate feedback to the student for test questions on the pretest, embedded tests, lesson or segment tests, and post-tests.
4.	Documentation of the student path through the ICW, and the time spent on specific lessons, segments, or topics.	Determination of student mastery of objectives.
5.	Disenrollment of students from the course.	Reporting of student performance information.

F-2. Tests in ICW. The types of tests usually developed in ICW courses are pretests and criterion tests. Use a pretest to measure the student’s ability to attain each objective, before developing ICW, and before entering students in an ICW lesson. Use a criterion test to measure the student’s attainment of the objectives, and the effectiveness of the ICW. To design ICW tests:

- a. Use a student’s pretest score to branch the student to “need to know” information. This reduces boredom by not forcing the student to review items they already know.
- b. Inform students how many questions are on the test, and the anticipated time for completion, to help students gauge the intensity of the test.
- c. Allow students to “back out” of taking a pretest, if they do not know the content. Forcing students to take a test, when they realize they do not know the content, can introduce unnecessary stress into a learning situation.
- d. Provide clear instructions for taking the test, including options for changing answers, to reduce the possibility of students making errors when they actually have mastered the objective.
- e. Provide a method for students to review completed tests. If students respond with a wrong answer, and subsequently realize it, this allows the student to correct the answer, just as in a paper-and-pencil testing situation.
- f. Provide immediate feedback to student answers, in the same order that they answer the questions, to reduce confusion, and increase the learning value of a test.

g. For remediation, design the program so the computer “works through” a problem (provides real-time help) interactively for students, instead of just providing the correct answer. Where a student has a partially correct answer, the computer identifies the point where the student is in error, and invites the student to continue from that point, thus reducing learning time.

F-3. Two primary methods to test learners using IMI.

a. Simulated/actual hands-on performance tests. This method requires learners to perform a simulation of a task (TLOs/ELOs), or actually perform those TLO/ELOs performed within the constraints of the IMI delivery system, to prove mastery. Note: This is the preferred method of testing via IMI. Use this method as much as the medium allows.

b. Knowledge-based (written test) options. Using this method, an assessment is made of the learner’s ability to apply facts, principles, procedures, etc., required to perform the LO. Knowledge-based tests commonly use sets of essay, short answer, matching, drag-and-drop, and multiple-choice questions. Note: The computer cannot grade essay tests; a grader is required to score these tests.

F-4. Design, development, and implementation procedures for IMI tests.

a. Use visual, audio, and graphic components and simulations, to develop the test as realistically as possible. Note: If the same personnel that develop the IMI do not develop the tests, merge storyboards prior to production, since testing and training use many of the same visuals. This increases—

- (1) Learner interest.
- (2) Transfer of learning.
- (3) Learner retention.

b. Develop a simulated performance test that provides a two or three-dimensional simulation of the required performance. Ensure the test developed simulates on-the-job environments as closely as possible; requiring the learner to complete the simulation, through application and synthesis of knowledge and skill presented in the instructional materials. This ensures the learner performs the objective to the stated, or required, standard. The performance asked on the test is a simulation, as close as possible to the job environment, to ensure the learner can perform the task in the actual task environment, under the actual condition and standards expected. The computer capability:

- (1) Randomly selects, and orders for delivery, the validated test items/sets that sufficiently measure each objective, based upon the test plan.

TRADOC Pam 350-70-5

(2) Randomly selects common equipment malfunctions, scenarios, and control and indicator settings; thereby allowing multiple validated versions of test/test items. Note: Take care when using random selection, to ensure that each item/set is sufficiently validated. When in doubt of the validity of the test item, do not use random selection. Only use properly administered, validated tests.

(3) Stores and gives feedback to the learners and trainers, based on learner response.

(4) Branches based on learner response(s) to retrain/provide remediation, only on those objectives needed.

(5) Collects and transmits test and test item analysis data. This will allow for the maximum utilization of the computer capability, both in areas of real time and task fidelity.

c. Include design techniques described in paras [F-2c](#), [F-2d](#), [F-2e](#), and [F-2f](#), above.

F-5. Design options. The designer determines the most effective and appropriate design options for the tested material. The tests may include equipment simulation, visual performance-oriented, and textual questions.

a. Simulated performance test. Provides a two or three-dimensional simulation of the required performance. This test simulates the on-the-job environment by requiring the trainee to complete the simulation, through application and synthesis of knowledge and skill presented in the instructional materials.

b. Discovery (heuristic) test. Presents the learner with problem-solving simulations that emulate an on-the-job environment, with stimulus information that is inadequate, incomplete, ambiguous, or irrelevant to the simulated environment. The learner is required to synthesize knowledge, and apply training received, in order to solve the problem(s) presented in the job performance simulation.

c. Simple gaming test. Presents test items in a gaming style.

F-6. Test preparation. Interactive Multimedia Instruction administers realistic performance-oriented, CRTs. These tests determine if training objectives were achieved, and measure what trainees actually know, or are able to perform, in relation to the standards of the training objective(s).

a. To prepare tests:

(1) Design or write the test, test item, or test item set.

(a) Randomly select items or conditions, as appropriate.

- (b) Provide learner feedback.
 - (c) Determine trainee's prior knowledge.
 - (d) Create storyboards and items to add.
- (2) Select graphics used in the course.
- (a) Rapidly display graphics.
 - (b) Add realism and interactivity to tests and exercises.
- b. Select the types of graphics for the IMI, based on the objective.
- (1) Equipment related uses a pictorial representation (video or computer graphics) of the actual equipment.
 - (2) Decisionmaking of soft skill uses a linear motion (video and audio) scenario, with appropriate courses of action from which the learner can select.
 - (3) Identification of items or forms uses a photograph, or drawing, of the item or form. Ensure that the photograph or drawing is clear, crisp, and free from distortion.

F-7. Delivery options.

- a. As part of the IMI design, there are two options for the delivery of the IMI test:
- (1) Test integrated within a module. The IMI tests are totally embedded within the instructional module, as an integral part of the total unit, including the objectives, material presentation, practices, test, remediation, and feedback. The actual "grading" of the posttest is accomplished within or outside of the actual module. See paragraph [F-9](#), below.
 - (2) Testing as a separate IMI module. This option designs and delivers the IMI test as a completely separate piece of IMI courseware, administered after the learner indicates their readiness to take the posttest. For optimal use, this requires the addition of a Learning Management System (LMS) to control the selection of the test version delivered, and the actual delivery to the learner workstation.
- b. For test control purposes, you may desire to separate the instructional presentation software from the testing courseware. Use the embedded test when test control is not an issue, and remediation is integrated for speed of delivery. The capabilities of the IMI authoring tool selected and the LMS might also dictate the selection of a particular delivery method over another. Technical consideration—such as high-quality, interactive, and/or distributed simulations—could dictate the delivery of testing of courseware as a separate item.

F-8. IMI test validation. Unique aspects associated with IMI test validation, and description of checks, to ensure the IMI test is administered properly, are addressed below.

- a. Since the learner—in most cases—is not given the opportunity to ask questions, ensure test instructions are absolutely clear, complete, and unambiguous.
- b. Ensure the learner can respond to the test item(s), as designed.
- c. Ensure all links, permissions, reviews, and navigation capabilities function as designed.
- d. Ensure graphics, video, and simulations used within the test are clear, and free of ambiguities.
- e. Ensure the learner has the capability to review/change responses, based upon the design of the test.
- f. Ensure learner responses are recorded/saved/transmitted, based upon the test design plan.

F-9. Test grading.

a. At the conclusion of testing, there are two options for grading, or scoring the test items and exams, based on the instructional design, and the programming capabilities of the software selected for the courseware. Use these two options, regardless of whether the test is an integral part of a total IMI courseware module, or administered as a separate module. Note: “Scoring” is the first step in determining a “grade” for a learner. Hereafter, the word “grading” is used to describe both scoring and the assigning of a PASS/FAIL “grade.”

(1) Test grading internal to the courseware: The IMI module or lesson itself, using the inherent capabilities of internal programming, may evaluate the learner response, “grade” the response, and provide feedback on the correct or incorrect response evaluated. Additionally, you may also design the IMI to further determine an overall score, compare the score with the GO/NO GO (passing cutoff) criteria, and provide learner feedback, as to whether the criteria was or was not met, that is, assign a “grade” of GO or NO GO.

(2) Test grading external to the courseware: Many IMI programs and lessons use an external LMS, which is called on to perform programmed “grading” functions. These functions may include the evaluation of raw learner responses; grading of the raw responses, or submission of the raw learner responses for grading; determination of an overall score; comparison of the learner score with the GO/NO GO criteria; and branching, which could provide feedback or remediation, as needed. Raw learner responses are sent from the IMI courseware module to the LMS for “grading.”

b. The nature of the test items frequently dictates which grading method is chosen to use. By its nature, the lesson cannot internally grade an item that requires manual grading. The courseware must deliver the learner product back to the LMS, for distribution to the instructor/grader, and ultimately deliver the response to the learner. The capabilities of the IMI authoring test software and related LMS play an integral part in the grading method chosen. If the test is graded electronically, consider how quickly grading and feedback is provided to the learner, as part of the grading method chosen. Due to the nature of communication links required, coupled with the risk of an interruption in the link itself, use of an LMS to provide grading may be a slower method.

F-10. Test feedback and remediation. At the conclusion of testing, provide the learner quality feedback and remediation that maximizes the learning experience. As soon as possible following the scoring of a learner test, give feedback and remediation at two levels—

a. Global. At the global level the learner should receive—

(1) Test score information.

(2) Subtest (section) score information, if applicable.

(3) Links (automatically or through learner interaction) to prescriptive or diagnostic training, for those areas where the established standards were not met (passing cut-off scores).

b. Test item. At the test item level, provide the learner—at a minimum—information on which items were missed, and the correct response to the items missed. The following features are desirable, but not essential:

(1) Hyperlinks to specific training material (e.g., technical manual page, table, or chart), which addresses those areas the test identifies as a deficiency.

(2) Remediation tailored to the learner's incorrect response(s).

F-11. Tests as learning objects and SCORM compliancy. Ensure ICW tests, designed under the Army's DLXXI IMI contract, and/or for delivery via the Army's LMS, and/or for storage in an Army DL-compliant learning object repository, are SCORM compliant. Design ICW tests as a separate Sharable Content Object that the learner can "open" when desired. The following references provide additional guidance:

a. See http://www.atssc.army.mil/itsd/imi/Documents/IMISOWTemplate_Jan03.doc for the latest Army SCORM IMI design specifications as attachments to the IMI Statements of Work.

TRADOC Pam 350-70-5

b. The Army's acceptance criteria for SCORM-compliant courseware are found at http://www.atsc.army.mil/itsd/imi/Accept_Criteria.asp.

c. The latest advanced DL SCORM specifications are found at <http://www.adlnet.org>.

Glossary

Section I

Abbreviations

APFT	Army Physical Fitness Test
ASAP	as soon as possible
CBT	computer-based training
CD	compact disk
CMI	computer managed instruction
CRT	criterion-referenced test
CTDPP	Course Test Development Project Plan
CTP	Course Testing Plan
DL	distributed learning
ELO	enabling learning objective
IAW	in accordance with
ICW	Interactive Courseware
IMI	Interactive Multimedia Instruction
IU	instructional unit
JA	job aid
kW	kilowatt
LAP	Learner Assessment Plan
LMS	Learning Management System

LO	learning objective
MOS	military occupational specialty
MSC	major subordinate command
NRT	norm-referenced test
QAE	Quality Assurance Element
QAO	Quality Assurance Office
QC	quality control
SAT	Systems Approach to Training
SCORM	Sharable Content Object Reference Model
SEP	Student Evaluation Plan
SME	subject matter expert
SOP	standard operating procedure
TD	training development
TDPMP	Training Development Project Management Plan
TDT	Test Development Team
TLO	terminal learning objective
TRADOC	United States Army Training and Doctrine Command
USAMPS	United States Army Military Police School

Section II
Terms

Cheating

The act of completing a form of assessment, outside the boundaries or rules set for the assessment. This includes submitting other's work as ones own, using information resources or tools the conditions of the assessment prohibit, or collaborating with other individuals in a manner the conditions of the assessment prohibit.

Checklist

- **Job aid:** Used to determine or ensure a process or procedure is followed. The execution sequence lists the elements of the activity. A check is usually placed beside each element as it is accomplished.
- **Test:** A list of actions identifying critical actions to perform that are objectively observed and measured, to determine student performance of the objective to the prescribed standard; the sequence of performance, if any; and identification of steps requiring accomplishment, if any. Actions are measured using the absolute measures GO or NO GO. The performer either performs or does not perform the action described in the learning objective; or, meets or does not meet the performance criteria.

Completion item

A test question calling for the completion (filling in) of a phrase, sentence, etc., from which one or more parts are omitted.

Conditions

One of the main parts of an objective that tells (1) what the student has to work with, (2) the environmental circumstances under which the performance is demonstrated, (3) what the student must work on, (4) the starting points, and (5) any limitations, special instructions, etc.

Correlation

Relationship between two scores or measures; tendency of one score to vary concomitantly with the other, as the tendency of high IQ students to have above average reading ability. The existence of a strong relationship (that is, a high correlation) between two variables does not necessarily indicate that one has any causal influence on the other. (See coefficient of correlation).

Criterion

The standard by which something is measured. In Army training, the task or learning objective standard is the measure of soldier/student performance. In test validation, it is the standard against which test instruments are correlated, to indicate the accuracy with which they predict human performance in some specific area. In evaluation, it is the measure used to determine the adequacy of a product, process, or behavior.

Criterion-referenced test (CRT)

A CRT measures an individual's skill or knowledge, compared to what the individual must perform or must know, in order to successfully perform a task. An individual's performance is compared to external criteria or performance standards, derived from an analysis of what is required to perform a particular task.

Difficulty index

A measure used for test item analysis, which indicates the percentage of examinees responding correctly to that test item.

Discrimination index

A measure used in item analysis, which compares the performance of masters to nonmasters.

Distributed learning

The delivery of standardized individual, collective, and self-development training, to soldiers and units, at the right place and time, through the application of multiple means and technologies. Distance learning may involve both synchronous and asynchronous student-instructor interaction, and may also involve self-paced instruction, without benefit of access to an instructor.

Distractor discrimination index

A measure used in item analysis of knowledge-based multiple choice test items, which computer calculates. It is similar to the discrimination index in that it uses the same master/nonmasters (upper/lower groups) differentiation, however, this value is computed for each distractor. An interpretation of the value identifies how many masters, and how many nonmasters selected each distractor. This information is needed so that each distractor is revised, as appropriate.

Distribution (frequently distribution)

A tabulation of scores from high to low, or low to high, showing the number of individuals that obtain each score, or fall in each score interval.

Entry skill test

A pretest designed to determine if a student already possesses a certain skill or knowledge needed as a prerequisite, before undertaking new instruction.

Entry test

A test containing items, based on the objectives, the intended students must master, in order to begin the course. According to Instructional Systems Development methodology, an entry test is used to ascertain whether the students possess the prerequisite skills to enter the course. Entry tests assess student competencies, and compare them to required student entry behavior. The US Army Military Police School (USAMPS) seldom uses entry tests.

Equivalent form

Any of two or more forms of a test, that are closely parallel with respect to the nature of the content, and the difficulty of the items included, and yields very similar average scores and measures of variability, for a given group.

Feedback

Information and data, provided both within and outside the training system, that indicates the efficiency or effectiveness of the system, or product. It is the data and information provided to the appropriate training proponent concerning the effectiveness and efficiency of the proponents training products. Also, information provided to a student concerning their training performance.

Fidelity

How well the actions, conditions, cues, and standards of a test item, or practical exercise, approximate those of the actual task in the field.

Hands-On Performance Measure

A type of performance measure that tests the individual on the apparatus for which they are trained. Example: A hands-on performance measure of generator repair requires the trainee to actually repair a generator. When used in this manner, refer to the hands-on performance test as a high fidelity test (needing only content and criterion validity). Concurrent validity is not affected, because the test itself is high fidelity.

Internal consistency

A method to determine reliability (frequently referred to as the split-half method). In this method, the odd items (#1,3,5,7, etc.) are correlated against the even items (#2,4,6,8, etc.). This correlation gives an indicator of the test's reliability.

Item analysis

The process of determining whether a test item functions as intended. Alternatively, the use of results on individual test items to determine effectiveness of the item. It is used to obtain feedback on training deficiencies, score exceptions, and improve future versions of the test.

Master(s)

Individuals competent at performing a given task. Masters can perform the task(s) for which they were trained. An anticipated master is a person, who by virtue of training and/or experience, is reasonably expected to pass the test. In effecting concurrent validity, "masters," however, really are using "anticipated masters," until they have passed our test.

Mastery

Completion and passing of a training segment the CRT was developed to test. This accomplishment indicates the trainee can perform at the minimal level necessary for successful task completion, or better.

Matching item

A test item calling for the correct association of each entry in one list, with an entry in a second list.

Mean

Adding up and dividing the total of all scores, by the number of scores, to calculate an arithmetic mean score.

Multiple-choice item

A test item in which the examinee's task is to choose the correct, or best, answer from several given answers, or options.

Nonmasters

Incompetent performers, those not knowledgeable in the subject matter tested, or those with inappropriate training. Persons, who by virtue of training and/or experience, are reasonably expected to not pass a test. Students not yet trained on the material in question are nonmasters of that material.

Normal distribution

A distribution of scores or measures that, in graphic form, has a distinctive bell-shaped appearance. In a normal distribution, scores or measures are distributed symmetrically about the mean, with as many cases at various distances above the mean, as at equal distances below it, and with cases concentrated near the average, and decreasing in frequency, the further one departs from the average, according to a precise mathematical equation. Assuming that mental and psychological characteristics are distributed normally is very useful in much test development work.

Norm-referenced test (NRT)

An approach to testing, in which an individual's test score is compared to the scores of other individuals, regardless of standards an objective specifies. A test that grades a student, based on the performance of other students taking the same test, and scored based upon relative standards, such as class standings, rather than upon absolute standards, such as job competency.

Objective test

A test that presents no possibility of difference of opinion among scorers, as to whether responses are scored right or wrong. It is contrasted with a "subjective" test (e.g., the usual essay examination to which different scorers may assign different scores, ratings, or grades).

Parallel forms

As used in reliability estimates, this technique tests a student population with parallel tests, which yield an estimate of test reliability. This term also refers to the preparation of two tests on the same material. Parallel tests are of equal length and equal difficulty, and do not contain duplicate test questions.

Percentile

(1) A point (score) in a distribution, below which falls the percent of cases indicated by the given percentile. Thus, the 15th percentile denotes the score, or point, below

TRADOC Pam 350-70-5

which 15 percent of the scores fall. "Percentile" has nothing to do with the percent of correct answers on an examinee's test.

(2) A value, on a scale of 100, that indicates the percent of a distribution that is equal to, or below it. For example, a score at the 95th percentile indicates an individual performed better than 95 out of 100 who took the test.

Percentile rank

The percent of scorers in a distribution, equal to, or lower than the score corresponding to the given rank.

Performance test

In contrast to a paper-and-pencil test, this test requires motor or manual response on the examinee's part, generally, but not always, involving manipulation of concrete equipment, or materials. "Performance test" is also used in another sense, to denote a test that is actually a work sample. This type may include paper-and-pencil tests (for example, a test in accountancy, taking shorthand, or proofreading), where the only materials required are paper and pencil, but where the test response is identical with the behavior about which information is desired.

Phi coefficient (ϕ)

A simple statistical technique for CRT item analysis, used if data is available that indicates (1) which people pass which items, and (2) which people are "masters" and "nonmasters."

$$\phi = \frac{AD - BC}{\sqrt{(A+B)(C+D)(A+C)(B+D)}} \quad \text{where:}$$

A= number of "masters" passing the item

B= number of "masters" failing the item

C=number of "nonmasters" passing the item

D= number of "nonmasters" failing the item

ϕ is also used as a measure of test-retest reliability, and concurrent or predictive validity. In such cases, the formula remains the same, but the letters refer to different measures.

Test Retest Reliability (1 st Administration of Test)			Concurrent or Predictive Validity (CRT Results)		
	Fail	Pass		Fail	Pass
Pass (2 nd administration of test)	B	A	Acceptable (Concurrent or predictive measure)	B	A
Fail	D	C	Unacceptable	D	C

Power test

A test intended to measure level of performance, rather than speed of response. This test imposes no time limit, or a very generous time limit. All Army knowledge-based tests should be power tests.

Practical constraints

Factors such as time, money, facilities, safety, personnel, supervision, etc., which impair administration of the test items, if conditions and standards remain as presently specified in an objective. For example, an objective requiring the firing of nuclear projectiles has practical constraints; which requires modifying the objective to substitute firing “dummy” nuclear projectiles in the test item.

Practice effect

The influence a previous experience with a test has on a later administration of the same test, or a similar test; usually, an increase in the score on the second testing, attributed to increased familiarity with the direction, kinds of questions, etc. Practice effect is greatest when the interval between testings is small, the materials in the two tests are very similar, and the initial test taking represents a relatively novel experience for the subjects.

Pretest

A test administered prior to presenting instructions in a course, to exempt students from the course, or from blocks of instruction within the course. This test should parallel the post-test. It is also an assessment which measures soldier or civilian task competency before training. As a measure performance against a criterion, results focus training on what soldiers/civilians need to know, and provide links to this prescriptive training. As a placement test, it allows for testing out of lessons, modules, or phases of a course. See "Performance test" and "Post-test."

Process measurement

Measurement of a process, rather than a product. Process measurement is indicated when an objective specifies a sequence of performances which are observed, and the performances are as important as the final product of the performances. It is also appropriate when product is not distinguishable from process, or when the product is not measurable for safety for other constraining reasons. Process measurement usually requires observing whether or not a performance is performed properly and/or

TRADOC Pam 350-70-5

quickly enough, and in the right sequence. An example of process measurement is scoring a GO or NO GO on a person's ability to properly execute an "about face" in drill and ceremonies.

Product measurement

Measurement of a product, rather than a process. Product measurement is appropriate if: (1) the objective specified a product, (2) the product is measurable in either presence or characteristics, and (3) the procedure leading to product can vary without affecting the product. An example of product measurement is observing a weapon to see if it is reassembled correctly. There is no need to watch the weapon as it is reassembled (the process), because the product is observed to see if it was reassembled correctly.

Rating scale

A measurement device in which a rater chooses a response from choices arranged in a continuum, such as from low to high, or good to bad. When using a rating scale for scoring, specify the rating a student needs to achieve criterion level for the performance the objective specifies. A rating scale could also assess entering behavior at the start of instruction. Rating scales usually show 3 to 9 points, representing levels of performance, from low to high.

Raw score

The first quantitative result obtained in scoring a test. Usually the number of right answers, number right minus some fraction of number wrong, time required for performance, number of errors, or similar direct, unconverted, uninterpreted measure.

Recall item

An item that requires the examinee to supply the correct answer from memory or recollection, as contrasted with a recognition item, which requires only identifying the correct answer (e.g., "Columbus discovered America in the year ____? ____" is a recall item, whereas "Columbus discovered America in (a) 1425 (b) 1492 (c) 1520 (d) 1546" is a recognition item).

Reclama

A formal student challenge to a test item, setting forth the reason for the challenge.

Recognition item

An item requiring the examinee to recognize, or select, the correct answer from among two or more given answers. See "Recall item."

Reliability

The extent to which the test/test item gives consistent results each time it is used. Examine a test item reliability anytime it is examined for validity. Reliability is a synonym for "consistency" or "repeatability." A test is considered reliable if it makes the same discriminations among individuals on multiple occasions.

Remedial instruction (or “remediation”)

Remedial instruction are those learning activities/events provided to a learner, after test grading, in preparation for retesting, to provide special, additional, or refresher training on those learning objectives not mastered.

Representative sample

A representative sample is one which reflects (represents) the population for which a test is intended. In order to try out test items on a representative sample, ensure the persons in the sample are similar to those for whom the test is intended. Thus, if a test is intended for people who completed basic combat training, compose a representative sample of people who completed basic combat training. If a test is intended for people who have completed a field wireman course, compose a representative sample of people who completed that course. If a population is sampled randomly, the resulting group is a representative sample of that population, and not of any other population.

Retest

A retest consists of a second or subsequent test administered to a learner, which covers the LO(s) not mastered (passed) on the preceding test administration.

Sensitive test material

Any learner measurement/testing material that, by its nature is controlled, to assure the validity of the test responses. Without this assurance, the goals of testing are not accomplished. Sensitive material may include, but is not limited to: individual test items, test booklets, test administration guides, adjunct test material, “scratchpads” and “notes,” checklists, and scoring keys used for performance and performance-based exercises and tests.

Simulation

A situation where phenomena likely to occur in actual performance are reproduced under test conditions, without using the real-life equipment. Simulation use complex simulators—a simulated helicopter is an example—or simple simulators (that is, a rubber bayonet).

Skills

The learned ability to perform a job related activity which contributes to the effective performance of a task performance step. While knowledge is often necessary for skills, the knowledge of how to perform an act is not the skill—the performance of the act is the skill. Riding a bicycle, for example, is a skill requiring performance of a related sequence of actions. A person may have knowledge of how to ride—tell how to sit, pedal, shift gears, brake, etc—without possessing the skill of riding.

Spiral development

The iterative process of continuously improving/updating a process, program, organization, or system, based on evaluation feedback. The change is made at the appropriate entry point, thus minimizing workload.

Standards

(1) The third main part of an objective, which specifies the criterion by which the action is evaluated (how well and/or quickly an action is accomplished). You may include several types of standards in any objective, any of which tell how well or how quickly the action is done. An objective may have both a standard of quality and speed.

(2) A statement which establishes a criteria for how well a task or learning objective is performed. The standard specifies how well, completely, or accurately a process is performed or product produced.

(3) The task standard reflects task performance requirements on the job.

(4) The learning objective standard reflects the standard that is achieved in the formal learning environment.

Subject matter expert (SME)

An individual with a thorough knowledge of a job (duties and tasks), who is well qualified in the subject matter, and is usually trained and experienced in a particular subject area. This knowledge qualifies the individual to assist in the training development process (that is, consultation, review, analysis, etc.). (Frequently used at USAMPS as a synonym of instructor.) Subject matter experts normally instruct in their area of expertise.

Systems Approach to Training (SAT)

The Army's training development process. It is a systematic, spiral approach to making collective, individual, and self-development training decisions for the total Army. It determines whether or not training is needed; what is trained; who gets the training; how, how well, and where the training is presented; and the training support/resources required to produce, distribute, implement, and evaluate those products. The process involves five training related phases: analysis, design, development, implementation, and evaluation. See "Training development (TD)."

Task

A clearly defined and measurable activity individuals and organizations accomplish. It is the lowest behavioral level in a job or unit that is performed for its own sake. It is specific; usually has a definite beginning and ending; may support, or is supported by other tasks; has only one action and, therefore, is described using only one verb; generally is performed in a relatively short time (however, there is no time limit, or there is a specific time limit); and it is observable and measurable. Ensure the task title contains an action verb and object; it may also contain a qualifier. Types:

- **Collective task**

A clearly defined, discrete, and measurable activity, action, or event (that is, task) which requires organized team or unit performance, and leads to accomplishment of a mission or function. A collective task is derived from unit missions or higher

level collective tasks. Task accomplishment requires performance of procedures composed of supporting collective or individual tasks. A collective task describes the exact performance a group must perform in the field under actual operational conditions.

- **Common task**

- **Common skill level task** - An individual task every soldier in a specific skill level performs, regardless of MOS or branch, e.g., a task all captains perform.
- **Common soldier task** - An individual task **all** soldiers perform, regardless of rank. (Example: All soldiers perform the task, "Perform mouth-to-mouth resuscitation.") Note: There are common soldier tasks that apply to all Army civilian employees as well, e.g., maintain security of classified information and material.

- **Critical task**

A collective or individual task a unit or individual **must** perform to accomplish their mission and duties, and survive in the full range of Army operations. Critical tasks are trained. Types of tasks that are identified as critical include collective tasks, common skill level tasks, common soldier tasks, individual tasks, and shared tasks.

- **Individual task**

The lowest behavioral level in a job or duty that is performed for its own sake. It should support a collective task; it usually supports another individual task. Individual tasks include common soldier tasks, leader tasks, common skill level tasks, and organizational level tasks.

- **Organizational level task**

Common skill level task shared by other skill levels, e.g., company captains and first sergeants may perform the same tasks.

- **Shared task**

- **Organizational** - See "Task: Organizational level task."
- **Shared individual task** - An individual task soldiers from different jobs and/or different skill or organizational levels perform. Shared tasks are usually identified when conducting an analysis of a specific job. (Example: The lieutenant and sergeant in the same platoon perform some of the same tasks.)
- **Shared collective task** - A shared collective task is a collective task that applies to, or is performed by, more than one type unit, e.g., to units which have different proponents, or to different echelon/table of organization and equipment units within a single proponent's authority. Since the task, conditions, standards, task steps, and performance measures of shared

collective tasks do not change, all units that “share” the task train and perform the collective task in the same way.

Task analysis

An analysis to determine the skills and knowledge necessary to perform a task (or tasks), equipment and/or facilities required, attitudes required, critical tasks, proper sequence of actions, etc. Sometimes all the tasks in a given job are analyzed by a procedure called “job task analysis” or “job analysis.” Often, task analysis is used as a synonym for job analysis.

Terminal learning objective (TLO)

The main objective of a lesson. It is the performance required of the student, to demonstrate competency in the material taught. A TLO describes exactly what the student is expected to perform, under the stated conditions, to the prescribed standard on lesson completion. There is only one TLO per lesson, regardless of presentation method or media, and it has only one verb. The TLO may cover one critical task, part of a critical task (that is, a skill or knowledge), or more than one critical task. The TLO may be identical to the critical task taught, or a disparity may exist between them. Where there is a disparity, it is the TLO standard that the student achieves to demonstrate competency for course completion. See "Learning objective (LO)" and "Enabling learning objective (ELO)."

Test

A means of examination, trial, or proof; a series of questions or problems designed to determine knowledge or competency. A device, technique, or measuring tool used to—

- Determine if a student or group can accomplish the objective to the established standard.
- Determine if training does what it is designed to do, efficiently and effectively.
- Measure the skill, knowledge, intelligence, abilities, or other aptitudes of an individual or group.
- Collect data, as a basis for assessing the degree that a system meets, exceeds, or fails to meet the technical or operational properties ascribed to the system.

See "Criterion-referenced test," "Knowledge-based test," "Norm-referenced test," "Performance test," "Performance-based test," "Placement test," "Pretest," and "Testing out."

Test control

The application of security measures to protect tests, test items, and related sensitive material from unauthorized disclosure, from the time of their creation, until they are obsolete or destroyed.

Test item

Synonymous with test question.

Test plan

A test plan states the when, where, what, and how tests are administered during a lesson, module, phase, and course. The most atomic unit of test plans is for a lesson. A lesson test plan consists of the when, where, what, and how each test in the lesson is administered. A test plan for a course is composed of the test plans for each of its subcomponents (that is, phase, module, and lesson). A test plan for a phase is composed of the test plans for each module and lesson in the phase; and the module test plan is composed of the test plans for each lesson in the module.

Test-retest reliability

Determination of the stability of test scores by repeated testing. Test-retest reliability assumes that no training or forgetting takes place between test administrations, so both administrations are given close together in time. If a test has high test-retest reliability, a person should score about the same each time the test is taken. If it has low test-retest reliability, a person's score may vary widely from one test administration to the next.

Test review

A test review is a learning event/activity that occurs after the grading of tests, which provides to the learner the items that were missed, and short immediate remediation on the correct answer. This remediation usually consists of identification of the correct answer, and the reasons their answer was wrong and/or the correct answer was correct. **Note:** To avoid confusion with a slightly different procedure, do not refer to a test review as an after-action review.

Type I/Type II Errors

These are evaluation errors and are also called "false negatives" and "false positive" errors. A false negative (type I) error occurs when a competent person that has mastered the task, is given a failing score. A false positive (type II) error occurs when a trainee is given a passing score, but is not really a master of the task.

Validation

A process that determines if training products and materials perform as intended; course/courseware complies with all applicable policy/guidance; and personnel receiving training can perform to the objectives to standard. An iterative process through which a course is revised until it is effective in realizing its instructional goal.

TRADOC Pam 350-70-5

FOR THE COMMANDER:

OFFICIAL:

ANTHONY R. JONES
Lieutenant General, U.S. Army
Deputy Commanding General/
Chief of Staff

/signed/
JANE F. MALISZEWSKI
Colonel, GS
Chief Information Officer