

Navy ILE Guidance on Assessment Development



Approved for public release; distribution is unlimited.

15 August 2007

Change Record

Paragraph	Description of Change	Date	Authorized By
Not Applicable	Changed title page graphic and distribution statement	10 Apr 07	J. Aplanalp
Not Applicable	Updated JTA data model and associated information throughout document. Updated ILE URL.	15 Aug 07	J. Aplanalp

****This version replaces MPT&ECIOSWIT-ILE-GUID-2A**

Acronyms, Abbreviations, Definitions

ADL	Advanced Distributed Learning
CAM	Content Aggregation Model
ID	Instructional Design
ISD	Instruction Systems Design
LCMS	Learning content management system
LMS	Learning management system
PQS	Personnel Qualification Standards
SME	Subject Matter Expert

See the ILE website (<https://www.netc.navy.mil/ile>) for a complete list of acronyms, abbreviations and definitions.

Table of Contents

1.	Purpose	1
2.	Policy.....	1
3.	Action	1
4.	Assessment Integrity	1
5.	Breach of Security or Compromise	1
6.	Life-Cycle Maintenance.....	2
7.	Technical Requirements	2
8.	Assessment Strategy	2
9.	Test Design	6
9.1.	Pretest	6
9.2.	Practice Questions and Quizzes	6
9.3.	Section and Lesson Progress Test	7
9.4.	Module Test.....	8
9.5.	Computer-Delivered Assessment.....	8
9.5.1.	Assigned Passing Score.....	8
9.5.2.	Duration	9
9.5.3.	Weights.....	9
9.5.4.	Number of Tries.....	9
9.5.5.	Mandatory Delivery.....	9
9.5.6.	Shuffle	10
9.5.7.	Number of Assessment Items	10
9.6.	Evaluating Tests and Test Questions.....	11

List of Tables

Table 1	Learning Outcomes and Assessment Types.....	4
Table 2	Types of Assessments and Purposes	5
Table 3	Progress Test Structure.....	7
Table 4	Module Test Structure	8
Table 5	Computer-Delivered Assessment Guidance	10

1. Purpose

The purpose of this document is to provide guidance for designing, developing, and implementing assessments within the confines of the ILE.

2. Policy

Decisions pertaining to how assessments will be designed, developed, and implemented within Component Commands are the responsibility of the respective Command Content Sponsors. The one ILE technical requirement is that the software application(s) selected must be approved for use by the ILE Content Management Team, which can be contacted at 850-452-1001 extension 1.

3. Action

All Component Command Content Sponsors shall first determine how assessments will be used to support the instructional goals of their education and training requirements. Second, the Content Sponsors will determine the appropriate delivery mechanisms (e.g., paper-based, computer-/browser-based) required to meet instructional goals. If an instructional or technical requirement cannot be met by the currently existing software applications, then the Content Sponsors will communicate the requirement to the ILE Content Management Team at 850-452-1001 extension 1 for assistance.

4. Assessment Integrity

Assuring that assessment items are stored within the ILE systems in the most appropriate manner to prevent breaches in security is a Content Sponsor responsibility. Content Sponsors should provide written policy to subordinate commands as to how assessment items/banks will be administered. If the results of a learner's performance on the assessment instrument will directly affect a learner's promotion potential and career progression, the assessment is considered "high stakes."

- All high-stakes assessment items will be stored within the ILE in a manner that ensures the items will be seen only by the learner while engaged in the intended high-stakes assessment instrument.
- All high-stakes assessment items will be adequately protected from compromise by ensuring that they can be accessed only by designated personnel. Preferably, high-stakes assessment items will be partitioned in a manner that maintains separation between high-stakes and lower-stakes items.
- Additionally, if at all possible, assessment items for high-stakes assessments will not be the same as or very similar in wording to those items residing in lower-stakes item banks.

5. Breach of Security or Compromise

Content Sponsors are to identify preventive measures that may be taken to eliminate compromise or breaches in security.

- The ILE systems afford commands a number of ways to reduce breaches in security. Non-technology-based solutions are encouraged as well (e.g., posting

Core Values in the testing labs; discussing cheating with learners; requiring learners to sign a document before taking the exam; requiring identification checks before entering the exam room; providing a proctor in the exam room; apprising learners of the consequences of test integrity violations, etc.).

- In the event of a test security breach or compromise, the command responsible for the test should take appropriate action as provided by Content Sponsor guidance. Any technology-based test security breach must be reported and described in sufficient detail to the ILE Content Management Team as soon as possible. In the event of a high-stakes breach, the ILE Content Management Team may elect to secure the bank until a resolution can be identified.

6. Life-Cycle Maintenance

Content Sponsors are to provide guidance to their subordinate commands regarding the periodic review of assessment items. The review should include a table that matches the objective, the assessment items, the content, and the associated reference publication. The reviewing team should check to ensure that the information contained within the assessment items reflects the instructional content and the learning objectives.

7. Technical Requirements

SCORM does not address assessment issues via the Content Aggregation Model (CAM). During the planning phase, the instructional and technical design teams will consider how assessment instruments will be aggregated and sequenced. Knowledge-based assessments delivered by Interactive Multimedia Instruction (IMI) mediums must be QTI version 1.2 Compliant (<http://imglobal.org/question/index.html>). Compliance with the QTI 1.2 specification will ensure standardization of assessments and facilitate reusability of the items.

Functional requirements for testing are addressed through the Run-time Environment of the learning management system (LMS) and/or the learning content management system (LCMS). At a minimum, Content Sponsors should collect data including the following:

- (1) Learner identification
- (2) Content tracking
- (3) Item analysis
- (4) Validity and reliability information for high-stakes exams
- (5) Learner reaction or satisfaction to the learning event
- (6) Number of attempts

Content Sponsors have the responsibility to determine and document viable assessment solutions.

8. Assessment Strategy

Assessment, in its broadest use, describes the processes and tools designed to solicit data from learners and to make inferences about what those learners know or can do.

Designing an assessment strategy is integral to offering effective education and training opportunities.

During the analysis stage of the instructional systems design (ISD) process, learning objectives and learning outcomes are identified.¹ Decisions about what to assess and how to assess are part of the initial design process. Content Sponsors should include Science of Learning Practitioners and Human Performance Technologists during the assessment design phase to ensure that current learning theories and performance requirements are supported by the assessment selections.

Assessments make inferences regarding what learners know and can do. During test design you will decide what to test, how to test, and when to test. This analysis will be based on the content of the learning objectives.

Will the student demonstrate knowledge by answering questions or by applying knowledge in a real-world or simulated setting with performance observed by experts? The assessment plan identifies the knowledge and behaviors (e.g., cognitive, motor, verbal, social, and affective/attitudinal) that must be attained and which indicators will best demonstrate attainment of the intended learning & performance outcomes. See Table 1 for a description of intended learning outcomes and associated assessment types.

¹ The creation and management of content follows five basic steps: analysis, design, development, implement, and evaluation. This generalized framework of the instructional design process is called ADDIE. Although ADDIE provides a general process, there are a number of ISD methodologies and models that may be used in the design of learning materials.

Table 1 Learning Outcomes and Assessment Types

Learning Outcomes	Assessment Types
Cognitive	Drag-and-Drop and Matching Essay Identify (Hotspot) Multiple Choice Oral Board Ordering Practicum Reporting Sample Work Product Scenario Short Answer (Fill in the Blank) Simulation
Motor	Case Study Checklist Demonstration Drill and Practice Essay Exercise Experiment Gaming Goal-Based Assessment Observation (demonstration, simulation, workplace) Oral Board Project Scenario Simulation
Affective/Attitudinal	Observation Presentation Questionnaire Reflective Writing Scenario Simulation

Performance standards provide information about the level of learning required. The standards specified in each learning objective will help to determine this requirement.

Some, or all, of the following questions may need to be addressed:

- To demonstrate attainment of the learning outcome, must a student correctly answer 80% of the questions on a test?
- Must a learner accurately describe each of the steps that must be taken?
- What actions must a learner take to satisfy learning objectives?
- What evidence demonstrates attainment of intended performance/learning outcome?

- Will the student demonstrate knowledge by answering questions or by applying knowledge in a real-world or simulated setting with performance observed by experts?

Some learning objectives will be identified as critical, while others may not. Those learning objectives that are identified as critical will need to be focused on to ensure mastery (see Table 2).

Table 2 provides a list of types of assessments and the purposes for their use.

Table 2 Types of Assessments and Purposes

When Administered	Purpose	Typical Examples	Typical Inferences Based on Performance
Prior to Instruction	Prescriptive or Diagnostic	Pre-Test	Does the student need an advanced organizer to aid in identifying what is important? Has the student already achieved the intended learning outcomes? Does the student have the prerequisite skills needed to begin the instruction?
During Instruction	Formative Progress	Practice Test Self-Assessment Quiz Module/Lesson Test	Is the student achieving the intended outcome? Is remediation needed? Where/when should remediation occur?
After Instruction is Completed	Performance Measure	Post-Test Exams PQS Board	To what extent has the student achieved the learning outcomes? Has the student met the expected standard? (criterion-referenced) How does the student rank relative to others? (norm-referenced) Has a learning intervention been effective?

Feedback is an important element in the learning process. The assessment strategy should include opportunities for learners to learn from their performance. Learners benefit by receiving feedback on their performance, guidance about how to improve, and training in self-assessment (Pellegrino, Chudowsky, & Glaser, 2001).

In addition to feedback, the assessment may direct the learners to remediation. Assessment results can identify knowledge or performance gaps in need of remediation. The instructional designer determines the type of remediation the learner will receive and the methodology. Remediation in the ILE may:

1. Direct the learner to additional instructional materials or learning experiences.
2. Instruct the learner to repeat certain portions of instruction.
3. Suggest equivalent or alternative methods of learning.

9. Test Design

All questions are designed for ILE delivery with reuse in mind. There is the possibility that an assessment may be delivered at several locations within a module, course, or lesson.

9.1. Pretest

Pretests can be used in the following ways:

- As an advanced organizer: The advanced organizer is a cognitive strategy that aids the learner in identifying the important portions of the instructional content so that attention is focused.
- To bypass training: This allows the learner to expedite training by passing mastered material. A pretest can be used for just-in-time training or refresher training in which time to learn is limited and may be tailored to the individual needs of the learner.
- To ensure mastery of prerequisite knowledge: This ensures that a learner has the knowledge and skills required to proceed with learning. If the learner does not have the prerequisite knowledge, feedback regarding mastery of previously taken training is provided and remediation may be required.²

9.2. Practice Questions and Quizzes

Practice questions and quizzes, also known as progress checks or "checks on learning", are presented to sample learner understanding and maintain learner interest and involvement. Normally these questions appear during or at the end of each enabling learning object and/or prior to beginning a new learning activity. Practice questions should provide the learner with immediate feedback, remediation, and the correct answer. The number of questions provided per section depends upon and should correlate with the content, and the length of the instruction being assessed to ensure

² The skill groups will identify specific knowledge required for a learning event. The specific knowledge may be prerequisite knowledge. It must be reflected in prerequisites for the section or lesson and may be tested to ensure that the learner may proceed with learning.

| learner mastery.

In interactive multimedia instruction (IMI), the learner should be provided with two attempts to answer the question or perform the action correctly.

Feedback states as a minimum:

- Correct Answer: "Correct," followed by an explanation of why the answer is correct.
- Incorrect Answer: "Incorrect." The screen indicates the learner's choice, places a "check" by the correct answer, and explains why the "corrected answer" is correct.

For Instructor Led Training (ILT), quizzes are typically used. The instructor provides feedback about the correctness of what the learners' have done and if incorrect, the nature of their mistake.

9.3. Section and Lesson Progress Test

Each section or lesson may conclude with a knowledge and/or performance-based progress test. The level at which a progress test happens will depend on the nature of the content covered. The number of questions provided per section depends upon and should correlate with the content, and the length of the instruction being assessed to ensure learner mastery. Practical exams require the completion of a procedure or process requiring hands-on psychomotor skills.

A knowledge-based progress test is a segment of assessment designed to determine learner comprehension of one terminal objective statement and several supporting objectives referred to as enabling objective statements. The following assessment hierarchy (Table 3) may exist within a progress test.

Table 3 Progress Test Structure

Progress Test Structure	Description
Progress Test (approximately 30 minutes)	Addresses comprehension of a terminal objective statement (and associated terminal learning object (TO)).
Progress Test Section (approximately 5 to 10 minutes)	Addresses comprehension of an enabling objective statement (and associated enabling learning object (EO)).
Test Question	Addresses comprehension of a facet of the enabling objective statement.
Progress Test Summary	Summary of learner performance and comprehension.

If a learner fails a section or lesson progress test, remediation and re-testing is as directed by an instructor or as directed by the LMS based on predetermined conditions. However, it is recommended that a progress test only be delivered to the learner two times. It is recommended that the passing score be at least 70% for a homogeneous

test. Some tests may require a passing score of 100% to ensure mastery. Policy on passing scores is set by Content Sponsors. Developers and Content Sponsors should consult with Science of Learning and Human Performance personnel for guidance.

9.4. Module Test

A knowledge-based module test is a segment of assessment designed to determine learner comprehension of multiple terminal objective statements and their respective enabling objective statement. The test may comprise about 50 questions. The following assessment hierarchy (Table 4) may exist within a module test.

Table 4 Module Test Structure

Module Test Structure	Description
Module Test (approximately one to two hours)	Addresses comprehension of multiple terminal objective statements (and associated terminal learning objects (TOs)).
Module Test Section (approximately 5 to 10 minutes)	Addresses comprehension of a terminal objective statement and respective enabling objective statements (and associated enabling learning objects (EOs)).
Test Questions	Addresses comprehension of enabling objective statements and respective facets of enabling objective statement.
Module Test Summary	Summary of learner performance and comprehension.

If a learner fails a module test, remediation and re-testing is as directed by an instructor or as directed by the LMS based on predetermined conditions. However, it is recommended that a module test only be delivered to the learner two times. The passing score should be at least 70% for a homogeneous test. Some tests may require a passing score of 100% to ensure mastery. Policy on passing scores is set by Content Sponsors. Developers and Content Sponsors should consult with Science of Learning Practitioners and Human Performance Technologists personnel for guidance.

9.5. Computer-Delivered Assessment

A knowledge-based test that is delivered via the computer from the LMS or a test management system may include certain conditions that affect the way the test is presented and scored. Determining how to construct assessment for computer-delivery will be based on several factors.

9.5.1. Assigned Passing Score

This is entered as a percentage of correct answers needed by the learner in order to pass the assessment. If the score is reached, the learner may proceed to new learning activities as directed by the LMS or instructor. Policy on passing scores is set by Content Sponsors. Developers and Content Sponsors should consult with Science of Learning Practitioners and Human Performance Technologists personnel for guidance.

- The passing score should be determined prior to question development and be based on the criteria defined in the learning objectives.
- The passing score should be able to distinguish critical differences in the level and type of comprehension obtained by learners.
- Consider providing information to the learner that identifies what a learner needs to learn to improve.
- In order to assign the passing score, consider reviewing the test questions to determine how many, and which ones, a learner must be able to answer correctly in order to meet the requirements of the learning objectives.

9.5.2. Duration

This indicates the length of time, in minutes, allotted to a learner to complete assessment.

- Consider the reliability of computer response if the learners are required to take assessment within a certain length of time.
- Only use a timed test if the criteria defined in the learning objectives require it.
- Consider the variations of the learner population in terms of their reading speed and hand-eye coordination.

9.5.3. Weights

This is a number that represents the value the assessment results will have in relation to other assessment scores.

1. The weight of a test question should be established to distinguish critical differences in the level and type of comprehension obtained by learners.
2. Higher weights should be assigned to test questions that provide strong evidence of learner understanding.

9.5.4. Number of Tries

This indicates the number of times a learner may retake assessment.

1. Consider allowing the learner to take assessment several times, if it is determined that the learner will improve comprehension by repetitive exposure to the same assessment.
2. The Content Sponsor determines the policy on the number of tries allowed.

9.5.5. Mandatory Delivery

This enforces that a test question is always included.

1. Mandatory delivery should be assigned to test questions that provide strong evidence of learner understanding.
2. Consider that learning can be enhanced by careful selection of questions to provide evidence of comprehension.
3. Consider that the careful combination of various test questions can provide strong evidence of comprehension.

9.5.6. Shuffle

This is used to rearrange test questions and test options.

1. Shuffle questions when the order in which the questions are presented is not of any significance.
2. Shuffle questions and distracters when you plan to reuse questions for several test types.
3. Consider a mix of question types when mastery of the learning objective is critical. Don't rely on the reliability of two-state questions.

9.5.7. Number of Assessment Items

This assigns the number of test questions that will be displayed to the learner during assessment delivery. This number may be equal to, or greater than, the number of mandatory items.

Deleted:

1. Set the number of test questions to less than what is available in the pool of questions when the number of questions presented to the learner is not of any significance.
2. Set the number of assessment questions to less than what is available when you plan to reuse questions for several test types. This will help to ensure that each delivery of assessment is unique.
3. Consider how many test questions a learner needs to answer correctly to be judged successful on a particular learning objective.

The computer-delivered assessment conditions are listed in Table 5 along with their application when used at the various levels within a test hierarchy.

Table 5 Computer-Delivered Assessment Guidance

Condition	When Applied at Test Level	When Applied at Test Section Level	When Applied at Test Question Level
Assigned Passing Score	Indicates the percent to pass test.	Indicates the percent to pass section.	Indicates the percent to pass test question.
Duration	Indicates the length of time allotted to complete test.	Indicates the length of time allotted to complete section.	Indicates the length of time allotted to complete test question.
Weights ³	Indicates the weight of test in relation to other tests.	Indicates the weight of test questions in a section.	Indicates the value of a test question option in relation to other question options.
Number of Tries	Indicates the number of times the test may be taken.	Indicates the number of times the section may be taken.	Indicates the number of times a test question may be taken.

³ Weights at any level can have an overall impact on performance scoring.

Condition	When Applied at Test Level	When Applied at Test Section Level	When Applied at Test Question Level
Mandatory Delivery	Indicates that the progress test is required to proceed.	Indicates that a section is required to proceed.	Indicates that the test question must be displayed each time the section is delivered.
Shuffle	Indicates random ordering of sections.	Indicates random ordering of sections.	Indicates random ordering of test question options.
Number of Assessment Items	Not applicable.	Indicates the number of test questions that are displayed each time the section is delivered even if a larger number of test questions exist within the section.	Not applicable.

9.6. Evaluating Tests and Test Questions

Tests, test questions, and test directions should be formatively evaluated before they are used to assess learner performance. The instructional designer (ID) and subject matter expert (SME) should ensure that the test directions are clear, that test questions are accurate, and that the conditions in which the learner performs the assessment are appropriately designed so that they address the requirements defined in the learning objective.

Prior to administering the test to the target audience, it may be necessary to preliminarily evaluate the test. A preliminary evaluation can be performed in the following way: set up a scenario in which a learner reads aloud to the ID the contents of the test, and explains what the learner must do. This kind of evaluation can reveal unclear directions, and confusing examples.

After a test is actually given, the ID and SME can assess the results for item clarity. Test questions that are frequently missed by most of the learners should be reviewed for inadequacies. The question may be too difficult, not covered in the instructional content, or confusing.

Test questions may be designed for reuse, and random display in IMI. When implementing computer-delivered assessment, there is a potential that the selection of questions delivered are not at the same level of difficulty as other selections. Therefore, the questions should be evaluated to ensure all test questions constructed for a learning objective are parallel and at about the same level of difficulty.

References

Donovan, M. S., Bransford, J. D., Pellegrino(Eds.) (2000). How People Learn: Bridging Research and Practice. The National Academy Press. Retrieved February 28, 2004, from <http://www.nap.edu/openbook/0309065364/html/R1.copyright> , 2000.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). Knowing what students know: The science and design of educational assessment. Washington, DC: The National Academy Press.