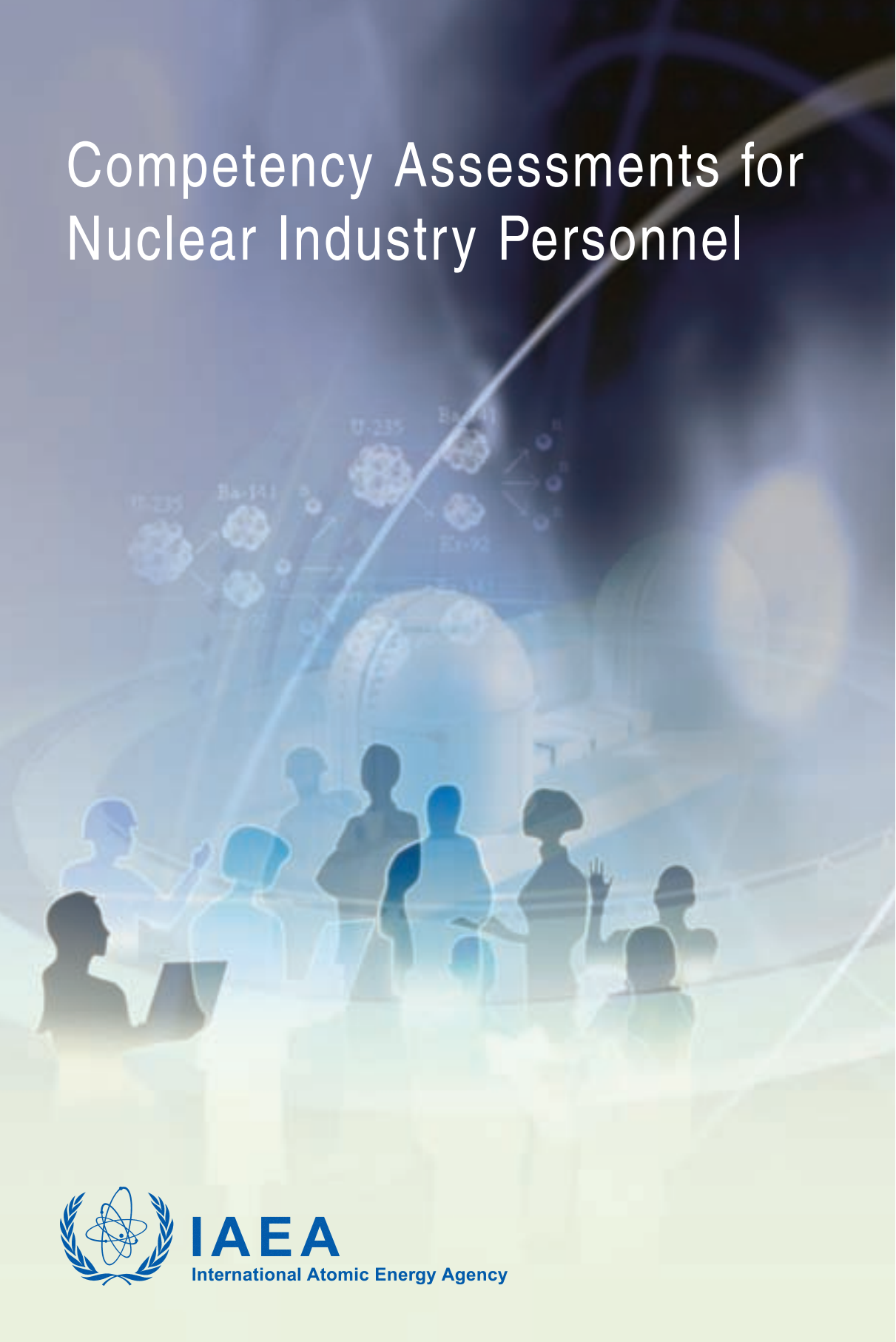
Competency Assessments for Nuclear Industry Personnel



IAEA
International Atomic Energy Agency

COMPETENCY ASSESSMENTS FOR NUCLEAR INDUSTRY PERSONNEL

The following States are Members of the International Atomic Energy Agency:

AFGHANISTAN
ALBANIA
ALGERIA
ANGOLA
ARGENTINA
ARMENIA
AUSTRALIA
AUSTRIA
AZERBAIJAN
BANGLADESH
BELARUS
BELGIUM
BENIN
BOLIVIA
BOSNIA AND HERZEGOVINA
BOTSWANA
BRAZIL
BULGARIA
BURKINA FASO
CAMEROON
CANADA
CENTRAL AFRICAN REPUBLIC
CHAD
CHILE
CHINA
COLOMBIA
COSTA RICA
CÔTE D'IVOIRE
CROATIA
CUBA
CYPRUS
CZECH REPUBLIC
DEMOCRATIC REPUBLIC OF THE CONGO
DENMARK
DOMINICAN REPUBLIC
ECUADOR
EGYPT
EL SALVADOR
ERITREA
ESTONIA
ETHIOPIA
FINLAND
FRANCE
GABON
GEORGIA
GERMANY
GHANA
GREECE
GUATEMALA
HAITI
HOLY SEE
HONDURAS
HUNGARY
ICELAND
INDIA
INDONESIA
IRAN, ISLAMIC REPUBLIC OF
IRAQ
IRELAND
ISRAEL
ITALY
JAMAICA
JAPAN
JORDAN
KAZAKHSTAN
KENYA
KOREA, REPUBLIC OF
KUWAIT
KYRGYZSTAN
LATVIA
LEBANON
LIBERIA
LIBYAN ARAB JAMAHIRIYA
LIECHTENSTEIN
LITHUANIA
LUXEMBOURG
MADAGASCAR
MALAYSIA
MALI
MALTA
MARSHALL ISLANDS
MAURITANIA
MAURITIUS
MEXICO
MONACO
MONGOLIA
MOROCCO
MYANMAR
NAMIBIA
NETHERLANDS
NEW ZEALAND
NICARAGUA
NIGER
NIGERIA
NORWAY
PAKISTAN
PANAMA
PARAGUAY
PERU
PHILIPPINES
POLAND
PORTUGAL
QATAR
REPUBLIC OF MOLDOVA
ROMANIA
RUSSIAN FEDERATION
SAUDI ARABIA
SENEGAL
SERBIA AND MONTENEGRO
SEYCHELLES
SIERRA LEONE
SINGAPORE
SLOVAKIA
SLOVENIA
SOUTH AFRICA
SPAIN
SRI LANKA
SUDAN
SWEDEN
SWITZERLAND
SYRIAN ARAB REPUBLIC
TAJIKISTAN
THAILAND
THE FORMER YUGOSLAV REPUBLIC OF MACEDONIA
TUNISIA
TURKEY
UGANDA
UKRAINE
UNITED ARAB EMIRATES
UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND
UNITED REPUBLIC OF TANZANIA
UNITED STATES OF AMERICA
URUGUAY
UZBEKISTAN
VENEZUELA
VIETNAM
YEMEN
ZAMBIA
ZIMBABWE

The Agency's Statute was approved on 23 October 1956 by the Conference on the Statute of the IAEA held at United Nations Headquarters, New York; it entered into force on 29 July 1957. The Headquarters of the Agency are situated in Vienna. Its principal objective is "to accelerate and enlarge the contribution of atomic energy to peace, health and prosperity throughout the world".

COMPETENCY ASSESSMENTS FOR NUCLEAR INDUSTRY PERSONNEL

INTERNATIONAL ATOMIC ENERGY AGENCY
VIENNA, 2006

COPYRIGHT NOTICE

Sales and Promotion Unit, Publishing Section
International Atomic Energy Agency
Wagramer Strasse 5
P.O. Box 100
A-1400 Vienna
Austria
fax: +43 1 2600 29302
tel.: +43 1 2600 22417
http://www.iaea.org/books

Printed by the IAEA in Austria
April 2006
STI/PUB/1236

IAEA Library Cataloguing in Publication Data

Competency assessments for nuclear industry personnel. — Vienna : International Atomic Energy Agency, 2006.
p. ; 24 cm. + CD-ROM.
STI/PUB/1236
ISBN 92–0–110105–8
Includes bibliographical references.

1. Nuclear industry — Employees — Evaluation. I. International Atomic Energy Agency.

IAEAL 06–00434

FOREWORD

In 1996, the IAEA published Technical Reports Series No. 380, Nuclear Power Plant Personnel Training and its Evaluation: A Guidebook. This publication provides guidance for the development, implementation and evaluation of training programmes for all nuclear power plant personnel using the systematic approach to training (SAT) methodology. The SAT methodology has since been adopted and used for the development and implementation of training programmes for all types of nuclear facility and activities in the nuclear industry. The IAEA Technical Working Group on Training and Qualification of Nuclear Power Plant Personnel recommended that an additional publication be prepared to provide further guidance concerning competency assessments used for measuring the knowledge, skills and attitudes of personnel as the result of training. This publication has been prepared in response to that recommendation.

A critical component of SAT (as part of the implementation phase) is the assessment of whether personnel have achieved the standards identified in the training objectives. The nuclear industry spends a significant amount of resources conducting competency assessments. Competency assessments are used for employee selection, trainee assessment, qualification, requalification and authorization (in some Member States the terminology may be 'certification' or 'licensing'), and job advancement and promotion. Ineffective testing methods and procedures, or inappropriate interpretation of test results, can have significant effects on both human performance and nuclear safety. Test development requires unique skills and, as with any skill, training and experience are needed to develop and improve them. Test item and examination development, use, interpretation of results and examination refinement, like all other aspects of SAT, should be part of an ongoing, systematic process.

This publication is primarily intended for use by personnel responsible for developing and administering tests. However, it should also be of value to line managers as well as to managers of training and human resources units. An understanding and appreciation of the importance of using valid and reliable methods for testing personnel will enhance the benefits and results of training, as well as ensuring that appropriate persons are selected for positions in the nuclear industry

The IAEA wishes to thank the participants and their Member States who contributed examples for this publication and for their valuable review and critique of it. Particular thanks are due to J. Yoder (United States of America) for the initial drafting and editing. The IAEA officer responsible for this publication was T. Mazour of the Division of Nuclear Power.

EDITORIAL NOTE

Although great care has been taken to maintain the accuracy of information contained in this publication, neither the IAEA nor its Member States assume any responsibility for consequences which may arise from its use.

The use of particular designations of countries or territories does not imply any judgement by the publisher, the IAEA, as to the legal status of such countries or territories, of their authorities and institutions or of the delimitation of their boundaries.

The mention of names of specific companies or products (whether or not indicated as registered) does not imply any intention to infringe proprietary rights, nor should it be construed as an endorsement or recommendation on the part of the IAEA.

CONTENTS

1. INTRODUCTION

1.1. OBJECTIVE

The purpose of this publication is to provide guidance on the development and implementation of competency assessments for nuclear industry personnel.

The concepts discussed are intended to be applicable and useful to all types of nuclear organization, including, but not limited to, nuclear power plants, research reactors, nuclear fuel cycle facilities, regulatory bodies and waste management facilities. However, the application of this guidance will vary depending upon the mission and structure of the organization and the complexity of associated jobs. This guidance is also applicable to the development and implementation of examinations for training in topical areas such as nuclear safety, radiation safety and fire protection, in other areas (for managerial, professional and technical staff) and for specialized technical training (e.g. non-destructive test examiners, radiographers).

This publication is primarily intended for use by personnel responsible for developing and administering tests. However, it should also be of value to line managers as well as managers of training and human resources units. An understanding and appreciation of the importance of using valid and reliable methods for testing personnel will enhance the benefits and results of training, as well as ensuring that appropriate individuals are selected for positions in the nuclear industry.

1.2. SCOPE

This publication focuses on competency assessments used for measuring the knowledge, skills and attitudes of personnel as the result of training. Other uses of competency assessments are briefly discussed in Section 2.

For some users, this book will provide a review of ideas and principles with which they are already familiar; for others it will present new concepts. While not intended to provide in-depth coverage of assessment theory, it should provide developers, instructors and assessors with a foundation on which to develop sound assessments.

1.3. RELEVANT IAEA PUBLICATIONS

This book provides supporting information for several other IAEA publications, principally: Recruitment, Qualification and Training of Personnel for Nuclear Power Plants [1], Nuclear Power Plant Personnel Training and its Evaluation: A Guidebook [2], Training the Staff of the Regulatory Body for Nuclear Facilities: A Competency Framework [3], Development of Instructors for Nuclear Power Plant Personnel Training [4], Assuring the Competence of Nuclear Power Plant Contractor Personnel [5] and Selection, Competency Development and Assessment of Nuclear Power Plant Managers [6].

While it provides guidance on performance examinations it does not provide details regarding assessment of nuclear power plant control room personnel with respect to their authorization or related simulator training assessments. The IAEA publication Use of Control Room Simulators for Training of Nuclear Power Plant Personnel [7] should be consulted for more in-depth guidance on these topics.

The available literature on test design provided by sources other than IAEA publications is very extensive (see Bibliography).

1.4. TERMINOLOGY

The terms related to examination development and implementation are provided in Appendix I. It should be noted that the terms 'assessment', 'examination' and 'test' are used interchangeably in this publication.

2. THE PURPOSES OF TESTING

The nuclear industry spends a significant amount of resources conducting competency assessments. Competency assessments are used for employee selection, trainee assessment, qualification, prequalification, authorization (in some Member States the terminology may be 'certification' or 'licensing') and job advancement/promotion. Ineffective testing methods and procedures, or inappropriate interpretation of test results, can have significant effects on both human performance and nuclear safety. Test development requires unique skills and, as with any skill, training and experience are needed to develop and improve them. Test item and examination development, use, interpretation of results and examination refinement, as with all other aspects of the systematic

approach to training, should be part of an ongoing, systematic process. Testing, and particularly the results of testing, can also be used for trainee motivation, instructional improvement and programme evaluation. In addition, testing can also be used to provide teaching and feedback. Discussion regarding each of the points above is provided in the following sections.

2.1. EMPLOYEE SELECTION AND JOB PLACEMENT

Tests are useful for trainee selection and job placement. Entrance tests are sometimes used as a basis for making the initial decision regarding the hiring of an applicant or the waiving of specific job training requirements after that individual begins work. In many Member States, ability or aptitude tests are used as a predictor of performance across a wide variety of jobs. In some Member States, tests designed to gauge personality traits such as honesty, integrity and attitudes are also used. Such tests provide assurance that the candidate being considered for a job possesses the prerequisite knowledge, skills and attitudes necessary to be able to complete the job specific training programme (if applicable) successfully, as well as be successful in the job. This ensures training resources are not wasted on candidates who are unfit for the job. Test scores may also indicate the need for remedial training. Many organizations also use aptitude tests for advanced placement of trainees in training programmes. Some use interest inventories, personality inventories and psychological evaluations to aid job placement. If an organization uses placement tests, the management, human resources department and training staff should make placement decisions based, in part, upon the results of these types of test.

2.2. TRAINEE ASSESSMENT

The purpose of a test should be to evaluate which competencies have been met by a trainee. Tests, when properly developed and conducted, should provide a valid and reliable indicator of learning objectives being achieved. Written, oral and performance examinations provide the most complete and efficient methods of collecting and documenting data on trainee competencies. Observations, interviews and other methods can also offer a significant amount of information.

2.3. QUALIFICATION, REQUALIFICATION AND AUTHORIZATION

These are very important assessments both from a safety significance perspective and also for the careers of those being assessed, as in most cases they cannot work in the industry without the appropriate qualifications/authorizations. These include both assessments used to grant initial authorizations/qualifications as well as periodic assessments conducted to reauthorize or requalify personnel. Such assessments often include a combination of written, oral and performance assessments.

2.4. JOB ADVANCEMENT/PROMOTION

Assessment centres (see Section 6.3) have become a common feature in a number of Member States and allow objective determinations to be made as to the person best suited for a job opening. Such objective assessments have the dual advantages of focusing the assessments on the criteria needed for success in the job, rather than on the candidate's success/capabilities in previous jobs having different requirements, as well as helping to avoid favouritism by selecting the best qualified candidate for the job rather than one already known. Assessments can also be used to identify long term career advancement paths.

In addition to assessing whether trainees have met the competencies required for a job or a task, the test results may also indicate that:

(a) The trainee qualifies for advanced placement or exemption from training;
(b) The material on a particular subject needs upgrading;
(c) A particular test question is poorly worded.

2.5. OTHER PURPOSES AND USES OF TESTING

2.5.1. Trainee motivation

Tests can be powerful motivators. Trainee study habits can be affected by test schedules. When tests are given on a daily or weekly basis, trainees may study more in anticipation of those tests. Likewise, when there is only an end-of-course test, many trainees may postpone studying until just before the test. Trainees are also generally motivated by the feedback that a test score provides. Low test scores could raise trainees' stress levels, which, if properly channelled, can result in increased concentration and study.

2.5.2. Instructional improvement

The results of testing can provide constructive feedback regarding the effectiveness of a training programme. Uniformly high scores for a topic or subject area may indicate that instruction was effective and can signify the readiness of the trainees for more detailed material or for the next step in the training programme. Conversely, low scores may indicate a need for improvement in instruction or teaching material, or that more instruction time is needed. For example, if a significant number of trainees missed questions based on one learning objective, training may be found to be inadequate for that objective.

2.5.3. Programme evaluation

Trainee test results can be combined and analysed to obtain training course or programme performance information. This information can be valuable in assessing programme strengths and weaknesses. To maximize the usefulness of the test data, systematic reviews should be conducted. Programmatic information can be obtained by analysing and interpreting the results of tests, and then comparing the data with information acquired from instructor, supervisor and trainee questionnaires. When combined, these sources can form a composite picture of programme strengths and weaknesses and appropriate actions can be taken to correct deficiencies.

2.5.4. An instrument to provide teaching and feedback

Testing is often viewed only as an assessment tool and not as a teaching or learning method. Recent research [8] has demonstrated that questioning can produce significant learning and performance benefits. An application of testing as a teaching method can be seen in the on-the-job training (OJT) process. In this activity, the trainees perform tasks under the supervision of a subject matter expert (SME). If the trainees perform properly, the performances are acknowledged; if not, the trainees are given immediate feedback on what errors were made and the proper steps needed to correct them. When this occurs during the training process, the trainees may have the opportunity to make the corrections at once. Testing can also provide effective feedback in the classroom, especially when test results are reviewed with the trainees. An open discussion of incorrect answers and why the wrong answers were selected can be very beneficial for both the trainee and the instructor.

The remainder of this publication is devoted to the application of competency testing to the development and implementation of examinations of

personnel in order to assess the results of training they received after being hired by nuclear industry organizations.

3. BASIS FOR EXAMINATIONS

Examination development should not be viewed as a strictly mechanical process with blind application of testing principles. Rather, in applying these principles, the test developer should be knowledgeable with regard to good testing principles, the subject matter, its significance and the most appropriate training setting and method for the material.

3.1. JOB/TASK ANALYSIS AND JOB COMPETENCY ANALYSIS

Specific areas to be tested should be those important to job performance. Proper analysis of the job or task for which the trainee is being trained provides direction for the entire training programme. In accordance with Ref. [2], job and task analysis or job competency analysis should be conducted. Competencies and tasks required for competent job performance are identified, documented and included in the training programme as a result of a job and task analysis or job competency analysis. Training objectives that identify training content and define satisfactory performance are derived from these tasks. Effective testing requires training objectives to be carefully selected and classified prior to test development. The establishment of a direct relationship between the actual job requirements, the training programme's learning objectives and the individual test items improves the validity and reliability of examinations (see Section 9).

3.2. TRAINING OBJECTIVES

Training objectives identify the knowledge, skills and attitudes that are necessary to perform the job or task. A properly designed training objective will allow the test developer to determine the specific aspects of the knowledge, skills, or attitudes to be measured by the test item. Training objectives also provide the conditions under which the test will take place and a standard against which test items are judged.

Along with training objectives, the test developer should review any available supporting instructional materials and other facility reference material to assist in test development. A more detailed discussion on training objectives is given in Ref. [2].

3.3. CLASSIFICATION OF TRAINING OBJECTIVES

Several models have been developed to classify training objectives into cognitive (knowledge), psychomotor (skills) and affective (attitudes) domains. Theses domains have been devised by Bloom and others [9–11]. The domains are referred to as learning taxonomies. It is important that test questions suitable for the class of training objectives are selected.

In the cognitive domain the ascending order of knowledge based mental processes are:

(a) Knowledge: Recognizes and recalls information.
(b) Comprehension: Interprets, translates or summarizes given information.
(c) Application: Uses information in a situation different from original learning context.
(d) Analysis: Separates wholes into parts until relationships are clear.
(e) Synthesis: Combines elements to form new entity from the original one.
(f) Evaluation: Involves acts of decision making based on criteria or rationale.

In the psychomotor domain the ascending order of physical skills are:

(a) Imitation: Observes a skill and tries to repeat it.
(b) Manipulation: Performs a skill according to instruction rather than observation.
(c) Precision: Reproduces a skill with accuracy, proportion and exactness. Usually performed independent of original source.
(d) Articulation: Combines one or more skills in sequence with harmony and consistency.
(e) Naturalization: Completes one or more skills with ease and becomes automatic.

In the affective domain the ascending order of attitudes (feelings, perceptions and values) are:

(a) Attending: Pays attention to received stimuli or events.
(b) Responding: Reacts positively to stimuli or events by participation.
(c) Valuing: Demonstrates belief in the worth or value of an event or activity.
(d) Organization of values: Compares various values and prioritizes them.
(e) Characterization by values: Displays an attitude characteristic of a pervasive, consistent and predictable set of values.

In the English language, the 'action' verbs used in a training objective or task statement can be used as an aid to determine the type of examination (written, oral and/or performance) and the format of test items (essay, multiple choice, etc.). Appendix II, Action Verbs, provides an example of a listing of action verbs suitable for the cognitive, affective and psychomotor domains and for the various levels within the domains.

For example, within the cognitive domain, the highest level is evaluation. Some of the action verbs associated with this level are: assess, explain and evaluate. In contrast, knowledge is the lowest level in the cognitive domain. Some of the action verbs associated with this level are: define, identify and label. For a training objective such as, "Identify the exposure limits for ionizing radiation" (lowest cognitive level), a suitable test method might be a written, multiple-choice examination. However, for a training objective such as, "Determine the reasons why personnel are violating plant safety rules" (highest cognitive level), such an examination method would not be suitable. An oral examination would be suitable in order to allow the examiner to probe the depth of understanding of the trainee.

3.4. SELECTION AND OVERVIEW OF ASSESSMENT METHODS

There is no single examination method appropriate for all situations. A method appropriate in one environment may be less appropriate in another. Each method has its advantages and disadvantages. Examination quality depends on the quality of the training objectives and the consistency between these objectives and the test items. Before examinations can be developed, the appropriate method should be selected. There are three basic methods: (1) written, (2) oral and (3) performance.

Written examinations are the primary type of examination used to assess knowledge and, to a lesser extent, attitudes. Written examination formats include multiple-choice, essay and short answer and other types of format such as matching and labelling questions. Oral examinations are the primary type of examination used to assess attitudes as well as knowledge. Oral examinations may range from interviews for managers and professional staff, to more structured one-on-one oral questioning and the use of oral examination boards

(for authorization examinations). Performance examinations are the primary means by which to assess skills but they can also be used to assess knowledge and attitudes, particularly with the use of simulators. Performance examinations are frequently combined with follow-up oral questioning. The type of performance examination may range from less formal assessments of satisfactory completion of job assignments for managers, professional/technical staff and regulatory staff to very structured on-the-job, laboratory and simulator examinations for operations, maintenance and technician personnel. Tables 1 and 2 illustrate the use of examination types versus the competence being assessed, their use for assessing the training of various job positions and the section of this publication which addresses the basic concepts and principles employed in the development and administration of the examination types.

In Section 4, the various formats used for written examinations are described, including their advantages and disadvantages. Table 3 contains a brief overview of the various formats.

Similarly, Sections 5 and 6 discuss the uses of oral examinations and performance examinations. Tables 4 and 5 provide a brief overview of some of the advantages and disadvantages of oral and performance assessments.

3.5. SUPPORTING MATERIAL

Appendices I–XV provide short examples of written, oral and performance assessment material. In addition, more comprehensive examples, some of which address all aspects of competency assessment, are

TABLE 1. COMPETENCIES BEING ASSESSED

Examination type	Competencies tested			
	Knowledge	Skills	Attitudes	Section(s)
Written	Primary* which should be designed to include higher cognitive levels	Secondary*	Secondary	4
Oral	Primary including higher cognitive levels	Secondary	Primary	5
Performance	Secondary	Primary	Primary	6

* Primary: the predominant method to assess the competency. Secondary: the supporting method to assess the competency.

TABLE 2. JOB POSITION TRAINING AND ASSESSMENTS

Job position	Types of training and assessment	Section(s)
Managers	Training is primarily by mentoring, coaching and soft skills training courses. Assessments are typically conducted using interviews (i.e. orals) and by completion of job assignments (i.e. performance) and selected use of written examinations.	Graded approach* to the application of Sections 5 and 6
Professional/technical staff (systems engineers, core physicists, etc.)	Training is also by mentoring, coaching and completion of job position specific assignments such as development of engineering design changes, procedure changes, calculations, modifications, etc. Technical training courses in systems, chemistry, nuclear physics, etc., are part of training provided depending on the engineering discipline. Assessments include written as well as oral questioning related to the satisfactory completion of structured job assignments.	Graded approach* to the application of Sections 4–6
Operations personnel	Training in theory and fundamentals, systems and components, and normal, abnormal and emergency procedures. Assessments include written, oral, and performance examinations in the facility and on a simulator and, where applicable, according to the type of nuclear facility or activity.	Full application of Sections 4–6
Maintenance personnel (electrical, I&C, mechanical) and technicians (health physics, chemistry)	Training in theory and fundamentals related to the job position with emphasis on hands-on training in a laboratory or OJT setting (using mentoring/coaching). Some use of written assessments with emphasis on the use of oral and performance examinations for task specific competencies (assessed on the job, mock-ups, laboratories or workshops).	Full application of Sections 4–6.

TABLE 2. JOB POSITION TRAINING AND ASSESSMENTS (cont.)

Job position	Types of training and assessment	Section(s)
Regulatory personnel: — Inspectors — Licensing and authorization personnel — Review and evaluation	Training for regulatory personnel ranges from classroom training to completion of job specific assignments under supervision (mentoring/coaching). Assessments may include written, oral questioning and completion of structured job assignments (particularly for task specific competencies).	Graded approach* to application of Sections 4–6
Instructors/trainers	Training for instructor competencies ranges from classroom training to completion of job specific assignments under supervision (mentoring/coaching). Assessments may include written, oral questioning, and completion of structured job assignments (particularly for task specific competencies).	Graded approach* to application of Sections 4–6

* To illustrate the graded approach, short examples have been included in the appendices of this publication. More extensive examples are included in the accompanying CD-ROM. These latter examples can also be found on the IAEA Electronic Nuclear Training Catalog (ENTRAC) on the internet at http://entrac.iaea.org. Registration is required, although there is no cost for access.

contained on the accompanying CD-ROM. The documents are also available on http://entrac.iaea.org

4. WRITTEN EXAMINATIONS

4.1. GENERAL DISCUSSION

Test items consist of two components: the content (what is asked) and the format (the way it is asked). The quality of the test item depends as much on the way the test item is constructed as on the content. Important topics that are tested by ambiguous, awkward, or poorly specified test items cannot be

TABLE 3. ADVANTAGES AND DISADVANTAGES OF WRITTEN TEXT FORMATS

Format	Advantages	Disadvantages
Short answer	Time saving related to developing and answering More definite than essay format 'Guess factor' is reduced Simple scoring	Not suitable for high level tests
Multiple-choice	Time saving related to answering and scoring No subjective scoring Scoring by mechanical means Definite (objective) answer	Guess factor is induced Difficult for development: — time consuming — finding appropriate distractors
Essay	Test higher level of knowledge, skills and attitude	Time consuming for: — scoring — answering

considered valid. Therefore, the selection of valid topics and the careful construction of test items are equally important parts of a valid test development process.

The conversion of a training objective into a test item is, in part, a creative process. Yet, there are certain procedures and guidelines that can help in writing a test item and in ensuring that the test item will measure the knowledge, skill, or ability that it is intended to measure.

Five steps are involved in developing test items:

(1) Test item format should be determined;
(2) Number of test items to be developed should be determined;
(3) Skill and knowledge test items should be written;
(4) Test item content should be validated;
(5) Test items should be incorporated into the test bank for future use.

Determination of test item format should be guided by the action verb of the training objective. However, any format of test item can be used to test any cognitive level if an appropriate stem (the introductory statement that functions to ask a question, define a task, or state a problem to be solved) is developed. Skill verbs suggest performance testing while knowledge verbs suggest one or more of the written formats. For example, 'startup' and 'shutdown' are skill verbs that suggest a performance test, whereas 'recall',

TABLE 4. ORAL ASSESSMENT

Type of oral	Advantages	Disadvantages
One-on-one oral	Assesses individual's mastery of training objectives for a given topic area. Allows probing the individual's knowledge by use of supplementary questions. Allows assessors to assess to some degree the individual's attitudes. Can be used for assessing qualification in a small topic area.	Labour intensive, requiring qualified SMEs to administer. Usually assesses only a small number of training objectives during each assessment owing to the time taken to administer. Requires some judgement of the assessor as to the correct responses. Not well suited for final authorization certification applications.
Oral board	Well suited for final authorization (certification) assessments as it involves more than one assessor. Can be extremely formal in nature. Can assess a large topic area of knowledge. Can be used to assess trainees' attitudes. Can be used to assess qualification in a large topic area.	Labour intensive to administer, requiring highly qualified (usually senior managers) SMEs to administer. Requires some judgement of the assessor as to the correct responses.
Progress review board	Used to determine trainee's mastery of training objectives while progressing through a training programme. Compares the trainee's progress with that of others in the training programme. Can be used as a predictor of trainee success in the training programme. Addresses the need for trainee remedial training. Can also be used as input to assess the training programme.	Limited application. Does not result in final qualification of the trainee.

TABLE 5. PERFORMANCE ASSESSMENT

Type of performance assessment	Advantages	Disadvantages
OJT	Conducted in a realistic setting (the workplace, or simulated workplace). Well suited for procedural-based performance training objectives. Formal in nature allowing use for staff qualification in topic areas. Variations of OJT assessment allow use for non-operations applications such as engineering and management training assessment and assessment of the mentoring process. Can be used to assess demonstrated understanding of attitudes such as procedural compliance, safety practices, self-check, etc. Reinforces management's expectations of the trainee's attitude. Although usually administered one-on-one, variations can allow for assessments of more than one trainee at a time.	Requires evaluator to be trained as qualified evaluator. Time consuming to administer. Can be resource intensive if the workplace is not accessible for assessment. Assesses a relatively small topic area. Limited to the number of trainees that can be monitored at one time.
Assessment centres	Good predictor of staff potential for future advancement. Results can be used as an input to succession planning. Can be used to identify training needs for current staff.	Has limited application. Difficult to develop.

'identify' and 'list' are verbs more appropriately addressed by the completion and short answer formats. Verbs such as 'discriminate' and 'select' should be tested using a multiple-choice format. Training objectives that require the trainee to classify or relate are most easily tested by using the matching format. Higher order verbs such as 'synthesize' and 'evaluate' may be tested using an

essay format. Appendix II provides a detailed listing of action verbs versus the domains and levels within the domains.

After the test item format is determined, the number of test items to be developed should be established. Some of the factors that significantly affect the number of test items to be developed are the number and type of examinations on which the test items will be used, the ease with which test items can be created and how frequently a test item will be reused. The initial number of test items selected for development should range from at least one to several items per learning objective.

Test items should be written after the format is selected and the number of test items has been determined. The items should have answer keys written at this time.

Content validation is the process by which test items are determined as being consistent between the content being learned and what is being tested. Each test item should be reviewed by SMEs. These validation reviews should be documented. Appendix III, Test Item Development Form, is an example of a test item development and review form.

A test item bank should be developed and maintained current. The test items in the test bank may be used to construct entry level, progress, or after instruction tests. A tracking system should be developed to correlate test items to the corresponding training objective.

The following sections offer guidance on how to select, construct and score different formats of written test items (e.g. short answer, multiple-choice, matching, essay). Traditionally, test items that require the trainee to supply an answer (e.g. short answer, essay) have been considered 'subjective'; test items requiring the trainee to select an answer (e.g. multiple-choice, matching) have been considered 'objective'. If graders require subject matter expertise to interpret the answers of test takers, the test item is considered subjective. If the examination can be scored without having to interpret the answer (e.g. machine scored), it is considered objective. An objective test item is defined here as one in which: (a) there is only one correct answer, and (b) all qualified graders would agree on the amount of credit allowed for any given trainee's answer. In a few Member States, multiple-choice questions may be used which have more than one correct answer. The use of such question formats requires special care and precautions.

Writing test items that are technically sound is a skillful and challenging task. Examination writers must balance technical concepts with an array of test item writing principles to create a single, 'good' test item. The test item should be one that reasonably assesses what a trainee ought to know. The test writing challenge is further complicated by the inherent ambiguities and interpretations of the written language. To make it perfectly clear in the mind of the

trainee what the test item is asking for, test items must be as precise and focused as possible.

Finally, the overall written examinations that utilize multiple-choice test items can be highly streamlined if examiners adopt the use of a separate answer sheet for the recording of trainee answers. The separate answer sheet affords several key benefits:

(a) Directions for recording answers for multiple-choice test items need only be given once (at the onset of the examination) rather than repeating 'circle the answer below' for each test item.
(b) Examiner scoring can be made easier and quicker by use of either an overlay key or an optical scanner.
(c) Test item analysis can be accomplished more readily through either a single sheet, visual profile of responses or by using an optical scanner to provide a computerized test item analysis.

In translating a training objective into a test item, the following examples may help the assessor generate ideas for potential test items. These examples not only help in the development of test items, but may also be useful when generating training objectives:

1. What are the common misconceptions about _____?

2. Why is _____ important to satisfactory job performance?

3. In what sort of circumstances might it be important to understand

   _____?

4. What might the individual do who does not understand _____?

5. What might the consequences be of a lack of knowledge about

   _____?

6. How can the individual demonstrate this knowledge?

4.2. DEVELOPMENT OF WRITTEN TEST ITEMS

Guidance on the construction of each test item format is provided in the following sections. However, there are basic principles that apply across all test item formats to:

(a) Ensure that the concept is relevant to the ability to perform the job;
(b) State the test item concisely;
(c) Choose the higher cognitive level;
(d) Make sure the test item matches the training objective;
(e) Omit unnecessarily difficult or irrelevant test items;
(f) Limit the test item to only one concept or topic;
(g) Avoid copying text directly from reference materials;
(h) Avoid 'backwards logic' test items (see examples);
(i) Place the easier test items at the beginning of each section;
(j) Ensure the test item can discriminate between those who have mastered the objective and those who have not.

Ensure that the concept being measured has a direct relationship to the ability to perform the job. The construction of the test item should clearly reflect the enabling objective. Word the test item so that it would be considered valid and reasonable to other SMEs using the same reference materials.

State the test item as concisely as possible, but provide all necessary information. The test item should be clear, grammatically correct and free of clues to the correct answer. It should be written at a reading level appropriate for the trainee. Often the individuals who develop a test item assume that certain conditions are inherent in the question when, in fact, they are not. It is important to have others review your test items to ensure that all necessary information is included, and that all excess information is deleted. You should ask yourself whether the trainees clearly know what they are expected to do. Do they have all the information they need to answer the test item? Does answering the test item depend on certain assumptions that must be stated?

When there is a choice between two cognitive levels, write your test item to reflect the higher level. Training objectives and test items should be written to reflect the level of cognitive domain that is most appropriate. The majority of the test items on an examination should consist of higher level cognitive items.

Make sure that the test item matches the training objective. It is very easy to end up with a test item that tests a relatively trivial aspect of an important training objective. When reviewing your draft test item, ask yourself whether it

is likely that someone could answer the test item correctly and still not meet the objective or perform the tasks.

Omit test items that are irrelevant. When reviewing your draft test item, ask yourself whether someone could do the job safely and effectively without being able to answer the test item. If so, is it because the content is inappropriate, the wording unclear, or the level of understanding too great?

Limit the test item to one concept or topic, unless a synthesis of concepts is being tested. Each individual test item should be reserved for testing one topic and that topic, as well as the intent of the test item, should be clear to both assessor and trainee. There is a common misconception that testing for multiple topics in one test item is a time efficient way to examine. Test items containing a variety of topics only serve to confuse the trainee about the purpose of the test item and, therefore, what is expected in terms of a correct response.

Avoid copying text directly from training or other reference material. Test items written in this way generally encourage rote memorization. Further, copying from reference material can cause confusion in test items because the material lifted often draws its meaning (and importance) from its surrounding context. Therefore, important assumptions or conditions stated elsewhere in the material are often omitted from the test item.

Avoid backwards logic test items — those test items that ask what should be provided in the test item — and provide what should be required in the trainee's response. It is important to test topics in a way consistent with how the topic should be remembered and used. For example, consider the following test item:

Test item:

If it takes 0.35 m^3 of concrete to build a square loading pad 0.15 m thick, what is the length of one side of the pad?

This test item gives the trainees information (the amount of concrete) they would normally be asked to calculate on the job, while it requires them to provide information (length) that would be supplied in an actual work situation.

In constructing your test items, make sure that you include information that trainees would typically have or have access to and require responses that reflect the decisions, or calculations, or other information they would typically have to supply. Place the easier test items at the beginning of each section. These test items help trainees gain composure and confidence. However, this is not to suggest that extremely easy test items be included in the examination for the sole sake of relieving trainee tension.

Finally, a test item must be worded so that it discriminates between those who have mastered the objective and those who have not. A well-written test item should parallel the objective that it is testing.

The following sections provide examples of short answer, multiple-choice, matching and essay questions. Appendix IV, Written Test Item Review Checklist, provides guidance for the review of written examination questions.

4.3. FORMATS OF TEST ITEMS

4.3.1. Short answer test items

Short answer items require the trainee to write a response in contrast to selecting from among a set of alternative responses. A short answer item may be presented in such a way as to give the trainee freedom to express ideas or it may restrict a response to a given content area or answer format. This test item format is useful in testing knowledge, comprehension and simple application. This format is more objective than essay format test items; however, it is not free from subjectivity in scoring. The guess factor is reduced from that of a multiple-choice test item because the chance of successfully guessing the answer is less. Because of varying trainee responses, the person grading the test item needs a definitive answer key and an understanding of the subject matter.

There are three common styles of short answer test items: (1) completion, (2) fill-in-the-blank and (3) list/diagram completion. In each case, the trainee must supply the correct answer. Because the developer of a short answer test item often has clear ideas about expected answers, it is essential that model answers, acceptable alternatives and associated grading criteria be written at the same time as the test item.

Examples of short answer test items are:

Completion:

How many protons and electrons does the element hydrogen contain?

Fill-in-the-blank:

A positive displacement pump should be started with its suction valve _____ and its discharge valve _____.

In addition to short answer and fill-in-the-blank question formats, drawing, labelling and diagramming formats are also sometimes used.

These formats are easy to construct and to score. However, the disadvantages frequently outweigh the advantages, since spelling errors, grammatical errors and legibility complicate the scoring process. Test graders cannot anticipate all possible responses and synonyms for the responses they expect and therefore often have to make subjective judgements on the correctness of trainees' responses.

Short answer test items can be used to elicit various abilities. These include:

(a) Recalling terminology, facts, symbols and classifications;
(b) Applying rules and making interpretations of data and other information;
(c) Solving scientific and mathematical problems;
(d) Manipulating symbols and balancing mathematical and chemical equations.

4.3.1.1. Short answer test item writing procedure

As with all test item formats, more consistent results are attainable if certain guidelines and rules are followed when the test items are developed. Keep the following in mind when constructing short answer test items:

(a) Provide clear, explicit directions;
(b) Avoid excessive wording or information;
(c) Ensure the answer key matches the test item;
(d) Avoid giving away the answer;
(e) Avoid trick (deceptive, clever, unusual, misleading) test items;
(f) Use the question form if possible;
(g) Word each test item in specific terms with clear meanings so the intended answer is the only one possible and so the answer is a single word, brief phrase, or number;
(h) Avoid copying statements word for word from texts or classroom materials;
(i) State the precision, numerical units, or degree of accuracy expected of the answer;
(j) Word the test item to avoid irrelevant clues or specific determiners;
(k) Word each test item so the blank lies towards the end of the sentence;
(l) Omit important rather than trivial words to be used in fill-in-the-blank test items;
(m) Use only one or two blanks in a complete sentence;
(n) If more than one blank is used, keep the blanks equal in length.

The test item directions should provide clear, explicit directions for answering the test item so that the trainee understands what constitutes a fully correct response. Choose words carefully to ensure that the conditions and requirements of the test item are appropriately conveyed. Words such as 'evaluate', 'outline' and 'explain' can invite considerable detail that is not necessarily relevant.

The test item developer should avoid excessive wording or information in the test item. In an attempt to make a test item operationally oriented and/or meaningful, there is a tendency to add more information than required for a correct response. The following example illustrates this point.

Two isotopes, one with a half-life of 10 days and another with a half-life of 5 days, sit on opposite trays of a set of scales. Assuming the isotopes have identical atomic weights and there is one curie of each isotope, which way will the scales tilt (which weighs more)?

The situation in the above test item bears little relationship to a situation that an operator would encounter on the job.

Make sure that the answer key response matches (and is limited to) the requirements posed in the test item. The test item below only asks for a yes or no response, yet the answer key indicates that an explanation of that one word response is required.

Test item:

Can a certified operator who has not been at the facility for a period of six months immediately resume duties in a licensed position?

Answer:

No. In any case the facility management would have to be notified before the operator resumes duties, and only after certification by an authorized representative of the facility or by a demonstration by the individual that his/her understanding of facility operations and administration is satisfactory.

The test developer should avoid giving away part or all of the answer by the way the test item is worded. In the test item below a half credit is allotted for answering the first part of the test item 'yes'. However, a test wise trainee can realize that the answer has to be yes, or else the second part of the test item would have read something like, "If so, how? If not, why not?"

Test item:

If the letdown line became obstructed, could boration of the plant be accomplished shortly after a reactor trip to put the plant in cold shutdown? (1 point) If so, how? (1 point)

Trick (deceptive, misleading) test items should be avoided. Trick test items can occur unintentionally when the answer key does not precisely match the test item. The test item below asks for how, not if, the termination criteria change.

Test item:

How do the SI termination criteria change following an SI re-initiation?

Answer:

They do not change.

4.3.2. Multiple-choice test items

Constructing a quality multiple-choice test item has two major advantages. First, the scoring of objective test items is considerably more reliable and less time consuming than scoring subjective test items. Second, since the test item requires less time to answer, more test items can be used to test trainee knowledge. This will provide better content coverage, which will also increase test validity.

There are two distinct parts of a multiple-choice test item, the stem and the choices, alternatives, responses, or options. The stem is the introductory statement and its functions are to ask a question, define a task, or state a problem to be solved. The choices, alternatives, responses, or options are the listed responses to the stem. The alternative part may be divided into two categories, the response that is the one correct or best answer and the remaining incorrect responses (termed distractors).

The multiple-choice test item can be a good selection to:

(a) Assess a trainee's ability to discriminate and make choices;
(b) Comprehend concepts, principles and generalizations;
(c) Make judgements about, and choices between, various courses of action;
(d) Infer and reason;
(e) Interpret new data or information;
(f) Apply information and knowledge.

The multiple-choice test item is not a good selection to use when the trainer needs to measure the trainee's ability:

(a) For recollection under minimum prompting;
(b) To explain and give examples;
(c) To produce and express unique or original ideas;
(d) To organize personal thoughts;
(e) To display thought processes or patterns of reasoning.

Multiple-choice test items require trainees to select the best answer from a series of choices. This series of choices can use calculations, descriptions, arguments, estimates, sketches, etc. They are useful for testing knowledge, comprehension and application of knowledge and theory. Since the answer is provided, the subjective aspect of scoring is eliminated. Furthermore, scoring can be conducted by mechanical means. If the stem is properly developed, multiple-choice test items can effectively test higher level, cognitive abilities, such as analysis, synthesis and evaluation. The latter abilities are difficult to measure directly through any type of written test item; however, the essay format of test items is suitable for this application.

The multiple-choice test item is the best choice where there are numerous test items covering a range of topics and the objectivity of scoring is the most important consideration, e.g. an examination that determines a trainee's selection for the job. The multiple-choice test item can be used to test a greater variety of training objectives and does not require the trainees to write out and elaborate their answers, minimizing the opportunity for less knowledgeable trainees to try to bluff or 'dress up' their answers to give the impression they know the answer.

Multiple-choice test items should generally not be used when:

(a) The test item calls for a numerical answer, as in computational problems;
(b) Writing the answer does not take any longer than using an answer sheet or marking the answer to the multiple-choice test item.

The development of multiple-choice test items from training objectives involves three steps:

(1) Create the stem of the test item by forming a question of an incomplete sentence that implies a question;
(2) Write the correct answer to the stem as consisely as possible;
(3) Write distractors that are plausible to trainees lacking the degree of knowledge you want the test item to assess.

TABLE 6. SUGGESTIONS FOR DEVELOPING THE STEM OF A MULTIPLE-CHOICE TEST ITEM

Do	Avoid
(1) If possible, write as a direct question. (2) If an incomplete sentence is used, be sure that: — It implies a direct question; — The alternatives come at the end (rather than in the middle) of a sentence. (3) Control the wording so that vocabulary and sentence structure are at a relatively low and non-technical level. (4) In test items that test definitions, place the word or term in the stem and use definitions or descriptions as alternatives.	(1) Using extraneous, superfluous and non-functioning words and phrases that are mere window dressing. (2) Using negatively worded test items. (3) Phrasing the test item so that the personal opinion of the trainee is an option. (4) Using textbook wording. (5) Cluing or linking test items (i.e. having the correct answer to one test item clued or linked to the correctness of the answer of a previous test item).

Table 6 provides suggestions for use in developing the stem of a multiple-choice test item.

4.3.2.1. Styles of multiple-choice test items (including matching test items)

Multiple-choice test items are the most common form of structured response-type test items. Multiple-choice includes alternate choice (true/false, yes/no, right/wrong) and matching. Both types vary only in the number of alternatives available.

Matching test items can be used where developers want to assess the ability to draw relationships between a similar set of ideas. While matching test items can be objectively scored, care must be taken to ensure that associations made in each column are similar in nature and kind (i.e. conditions and procedures, causes and results, actions and individuals).

The most common multiple-choice form consists of a stem and the choices. The choices include the correct answer and several distractors. This form is illustrated in the following example:

Test item:

The atoms of a chemical element all have the same:	STEM
(a) Atomic number (correct choice)	C
	H
(b) Relative atomic mass (distractor)	O
	I
(c) Number of neutrons in the nucleus (distractor)	C
	E
(d) Number of nucleons in the nucleus (distractor)	S

There are various ways of producing and selecting distractors. The test item developer can independently (or in discussion with colleagues) determine possible distractors. An effective method is to develop a simple test item of a fill-in-the-blank statement, e.g. "What do all the atoms of a chemical element have in common?" Then, give this test item to a group of trainees. Possible distractors that will emerge from their responses are usually more effective than those devised by the test item developer, particularly since the distractors also identify trainee learning errors.

The selection of distractors should be based on the required depth of trainee understanding. One set of distractors may be appropriate only when a general understanding of a subject is required, while a different set of distractors is needed when an in-depth understanding is required (see the examples in Appendix V).

Developing a good multiple-choice test item, written at the application level of the cognitive domain, which also adheres to good test item construction guidelines (i.e. valid, operationally oriented, plausible distractors), presents a challenge to the developer. The following models are presented as examples:

Model A:

(a) Correct answer
(b) Incorrect answer
(c) Incorrect answer
(d) Incorrect answer

Model A depicts the traditional multiple-choice design format. This model shows one correct single word/phrase answer followed by three incorrect single word/phrase options. The length of all options should be similar.

Model B:

(a) Correct answer
(b) Plausible misconception
(c) Incorrect answer
(d) Incorrect answer

Model B is a variation of Model A where a plausible misconception is used as an incorrect answer. Again the length of all options should be similar.

Model C:

(a) Correct answer with correct condition (e.g. because, since, when, if)
(b) Correct answer with incorrect condition
(c) Incorrect answer with correct condition
(d) Incorrect answer with incorrect condition

Model C depicts an acceptable multiple-choice design that uses conditions with answers (a condition in a setting, event, cause/effect that may make the answer correct or incorrect). Notice that Model C shows only one correct answer with its correct condition.

Model D:

(a) Correct answer
(b) Incorrect answer
(c) Correct answer with incorrect condition
(d) Incorrect answer with incorrect condition

Model D illustrates an acceptable model when it may not be possible to create all options in uniform length. This model shows paired lengths — two long and two short options — which avoids setting any single option apart (either too long or too short) from the remaining options.

The four models presented are basic. Other models may be developed and used in combination with one another.

In exceptional cases the use of a model with more than one correct answer may be used. For example:

Model E:

(a) Correct answer
(b) Incorrect answer
(c) Correct answer
(d) Incorrect answer

The use of this model has been questioned by some experts in the field of assessment, so it should be used with caution.

Appendix V contains examples of four option multiple-choice test items.

4.3.2.2. Multiple-choice test item writing procedure

Although multiple-choice test items are not as easy to construct as other forms, they are very versatile, can be used to test all levels and types of knowledge and minimize the likelihood of the trainee obtaining the correct answer by guessing. The following guidelines should be considered when constructing multiple-choice test items:

(a) Use four answer options;
(b) Do not use 'none of the above' or 'all of the above';
(c) Do not present a collection of true/false statements as a multiple-choice test item;
(d) Define the test item, task, or problem in the stem;
(e) Avoid negative stems;
(f) Provide sufficient counterbalance in the distractors;
(g) Include common misconceptions as distractors;
(h) Make answers plausible and avoid overlapping answers;
(i) Ensure that the distractors grammatically follow the stem.

The four answer multiple-choice test item is the most common and is the style preferred. The five answer option creates confusion for the test taker and any format with fewer than four answers makes guessing correctly more probable. Appendix IV contains examples for the development of multiple-choice test items.

4.3.2.3. Matching test item writing procedure

With matching test items, a trainee is required to match each word, sentence, or phrase in one column with a word, sentence, or phrase in another column. The test items in the first column are termed 'premises'. The answers

in the second column are termed ‘responses’. Consider the following guidelines when constructing matching test items.

(a) Give specific directions for each matching test item that indicate on exactly what basis the matching is to be done.
(b) State in the directions whether a response can be used only once or more than once.
(c) Use only logically related material in a single matching test item — nothing should be listed in cross-related columns that is not part of the subject of the test item.
(d) Consider the following possibilities for pairing premises and responses:

Premises	**Responses**
Terms or words	Definitions
Short test items	Answers
Symbols	Proper names
Causes	Effects
Principles	Situations in which they apply

(e) For any single matching test item, use only one set of premise–response formats (e.g. only terms with definitions).
(f) Each response should be a plausible answer for each premise, for example:

Test item:

Match the alarm (letter) with its appropriate tone (number). Responses (numbers) may be used more than once:

Premises	**Responses**
(a) Fire	(1) Pulse tone
(b) Reactor building evacuation	(2) Wailing siren
(c) Site evacuation	(3) Steady tone

(g) Present an unequal number of premises and responses, or allow responses to be used more than once — this will inhibit the trainee from obtaining clues to a correct answer through a process of elimination.
(h) Arrange the responses in a logical order, such as alphabetical or numerical.
(i) The entire matching test item should contain no more than six premises.
(j) Place the entire matching test item on one page so that the trainee need not flip back and forth between pages.

Directions to matching test items should not require more than one response to a premise. This practice not only fragments scoring and alters the implicit equality of the premises, but may also encourage the writing in of all responses to a premise in order to receive some partial credit (referred to as 'shotgunning').

If partial credit is possible, establish how it will be assigned in the examination key. For example, a five premise matching test item might have a 20% credit for each correct match.

4.3.3. Essay format test items

Test items that are termed essay format include those requiring the trainee to provide calculations, graphs, descriptions, arguments, estimates, sketches, or combinations of these. The trainee has to determine the answer and select the means of communication. Essay format test items should be used when the need is to have the trainee express himself/herself or to recall information from memory rather than to present the information to the trainee for recognition or identification. Test items of this format can test the whole range of abilities, but they are usually considered most suitable for application, analysis, synthesis and evaluation. The test item and the associated grading criteria must be prepared concurrently. The assessor grading the test item must be knowledgeable in the subject matter.

4.3.3.1. Styles of essay format test items

There are two major styles of essay format test items: structured and unstructured. In each case, trainees must supply the correct answer. However, the unstructured test items require a greater amount of analysis and composition than are required by structured test items.

The unstructured essay format test item is appropriate where a training objective expects a trainee to organize ideas, to develop a logical argument, to present evaluations of certain thoughts, to communicate thoughts and feelings, or to demonstrate other abilities requiring original written expression. An example of this format would be an assignment for the trainee to "Write a paper explaining the pros and cons of using written procedures." A disadvantage of the unstructured essay format test item is that scoring the response in an objective manner is difficult and time-consuming.

The structured essay format test item does not have to be limited to testing for recall and comprehension. It can be designed to require the trainees to apply their knowledge and skills to solve new problems or to analyse a novel situation. One way to do this is to write test item statements that require the

trainee to apply their knowledge to specific material given in the test. For instance, a description of a particular situation, along with the initial facility conditions, can be given in the test item statement. One advantage of the structured essay format test item is that by narrowing the focus of the test item statement to elicit specific and well-defined behaviours, the trainees are more likely to interpret the statement as the test developer intended. The structured essay format test item also prompts the test developer to provide more clarity in presenting the correct answers, which can lead to improved reliability of scoring.

A disadvantage of all essay format test items lies in their application, since they are often easier to develop than the multiple-choice test item. A good application of the essay format test item is to offer the trainee an opportunity to demonstrate abilities for written production, organization and expression and to identify interrelationships among ideas. A poor use of the essay format test item is to test for factual information. The trainer can make a more appropriate test item format selection by analysing whether the related objectives require the trainee to recall and supply answers or to recognize already supplied answers. If the latter applies, the multiple-choice test items are preferable because they allow for more objective scoring.

4.3.3.2. Examples of essay format test items

It is possible to express the same test item as either a structured or unstructured test item. Two extremes are shown here:

Structured essay format test item:

A pressurized water reactor at full power has the following conditions:

— Core flow (100%)
— T_{avg} (309°C)
— Core T (15.5°C)
— Steam generator T (T_{avg}-T_{steam}) (6.7°C).

The reactor is operating at steady state power with four loops operating. One reactor coolant pump trips, resulting in a three-loop operation. The rod control system is in automatic. The reactor has not tripped.

(a) Sketch graphs of RCS mass flow, reactor power, core T and steam generator pressure versus time if no operator action is taken. Use the same time axis for all graphs. (4 points)

(b) Calculate the final reactor power, T_{avg}, core T and steam generator T. (4 points)

This test item gives the trainee clear guidelines on what is expected. This is termed a structured test item and requires little personal initiative, but a clear requirement exists to apply knowledge gained from theoretical and practical experience in the course. Also, this test item lends itself to more consistent grading and is, therefore, more objective.

Unstructured essay format test item:

A pressurized water reactor is operating at 100% steady state power. One reactor coolant pump trips resulting in a three-loop operation. Explain what happens if the reactor does not trip and no operator action is taken. State any assumptions you make. (8 points)

This test item requires the trainee to interpret the word 'explain'. The trainee must decide what the important issues are, what assumptions to make (possibly based on parameters taken from his/her own experience) and then communicate these issues and assumptions by whatever means he/she thinks appropriate. Some trainees may use graphs and a short explanation; others may use only an explanation. Each answer will be unique. This test item requires the trainee to express clearly their understanding of the topic.

Unstructured essay formats are more difficult to grade consistently because of the variations that are possible, relative to the assumptions the trainee must make. Consequently, the grading becomes much more subjective in test items, as is illustrated in the foregoing example.

For assessment purposes, most test item developers prefer to write structured essay format test items. Unstructured essay format test items are most appropriate in learning exercises in which feedback and additional clarification can be provided if necessary (oral examination, OJT, walkthrough, etc.).

4.3.3.3. Essay format test item writing procedure

The layout of the test item and the clarity of the wording are crucial for the trainee to understand what is required when essay format questions are used. Generally the words 'how' and 'why' are the best, followed by 'what' and 'when', although the latter two can more easily lead to a single word or very short answer.

Trainees find particular difficulty with interpretation of words such as 'discuss', 'compare' and 'explain'. The level of answer expected is important in

determining whether a more specific statement is useful. The test item developer should follow a systematic method, such as the following:

(a) Define the behaviour the trainee is expected to exhibit or describe the process to be exhibited before beginning to write the essay format test item.
(b) Allow adequate time for developing the test item, model answer and grading criteria. Essay format test items usually require several revisions during the development process. Consider stating the anticipated answering time for the test item.
(c) Referring to the training objectives, select a topic from which the test item is structured. Ask only for important, relevant information.
(d) Ensure the test item is within the educational maturity level of the trainees.
(e) Determine the level of learning expected in response to a test item on the topic from the training objective. Use the words 'discuss', 'explain', or 'compare' only when deliberately aiming at higher cognitive levels.
(f) Define the problem explicitly — an essay format test item is of no use if the trainee cannot comprehend the test item. It must be open to only one interpretation so that all trainees address the same topic.
(g) Ask test items that require the trainee to demonstrate the ability to use essential knowledge and to do so in situations that are new or novel for the trainee, rather than simply recalling information.
(h) Limit the problem. The scope of a content area should not be too large. Remember that trainees will take longer to answer the test item than will the developer.
(i) Word the test item so the trainees know the limits of the tasks and their purposes and can answer them in the time allotted.
(j) Ask questions that are relatively specific or focused and which require relatively brief responses.
(k) If additional guidance for a particular test item is necessary beyond that shown in the general instruction, make a clear and concise statement.
(l) The SMEs should also agree on the relevance or importance of a particular test item, as well as the correct response.
(m) Develop the test item grading criteria. Identify the key responses expected from the trainee. Identify each necessary response (e.g. by underlining key words, phrases, or steps, or by circling important parts of graphs or diagrams). Determine the percentage value to be allocated for each test item subpart (based on their relative importance).

(n) Check the test item against the grading criteria:
  (i) Does the test item clearly request what is considered important in the grading criteria?
  (ii) Is the time needed to answer the test item proportional to the importance of the test item? Critical objectives justify more time if required.
  (iii) Do trivial details (e.g. an extended calculation) take too much time to address?
  (iv) How could the test item and/or the grading criteria be revised to be more consistent?
(o) Ensure that the sequence of requirements is in a logical order.
(p) Ensure that all the data (stimulus material) needed by the trainee are provided and clearly displayed together.
(q) Avoid test items that can be answered by a single word or short phrase.
(r) Avoid negative statements in test items; they may lead to double negatives that are difficult to interpret.
(s) State clearly whether exact answers or estimates (precise figures or sketches) are required.
(t) Ensure that the point value for each test item or part thereof is stated.
(u) Write the test item on a form such as that shown in Appendix III, Test Item Development Form.
(v) Word the test item so that the trainee can judge the approximate length of the answer desired and knows the point values or weight each will be given.
(w) Have another test item developer review the test item for accuracy, format and readability.
(x) Word the test item so that SMEs can agree on the correctness of a trainee's response.

4.4. OPEN REFERENCE TESTING

Training environments use several teaching aids that promote trainee learning. Many of these aids (e.g. textbooks, manuals and student guides) assist the trainee in mastering the required subject matter during the initial training phase. The trainee is expected to comprehend the material and to commit important facts to memory. However, there are many references (e.g. tables, charts, schematics and procedures) that trainees do not need to commit to memory (and should not) but instead have to interpret while performing their job.

To test trainee competence in the use of this material, test items that require the use of references to solve a problem or to reach a conclusion should

be developed. This requires that the reference, or a sufficient subset of the reference, be provided to the trainee during administration of the test. This use of reference materials during testing is referred to as open reference or open book testing.

The test developer should determine which references and their applications are necessary after reviewing the training objectives and the test specifications. While the open reference test is essentially no different than other written tests, there are several points to consider when using this method:

(a) References are considered tools trainees use to solve a problem or to reach a conclusion. The test items should not directly test knowledge of the references (this should be done in a 'closed book' section); rather, it should test for proper use and application of the references.
(b) Open reference test items should not be a 'direct look up' in a reference. For example, asking the specific set point for a protective action or the power of a motor are direct look ups. Rather, the reference material should be an aid in going through a series of knowledge based decisions in order to arrive at a logical conclusion or answer.
(c) Open reference test items should test the trainee's ability to locate, use and apply the information found in the references.
(d) Test items should postulate unique or varied circumstances that the trainees have not previously encountered. This makes the test item a true indicator of the trainee's ability to apply knowledge through the use of the references versus merely remembering an application from an earlier training session. Familiarity with routine applications of a reference can lower the learning level to simple recall.
(e) If a test item has references provided, it is not appropriate to make the test item more or less difficult than a comparable closed book test item. While very difficult test items may be useful in differentiating the most able trainees, they are not appropriate for job or task qualification. Do not make an open reference test item easier or more difficult than other formats.
(f) Keep test item requirements as close to real life situations as possible. It is important the test items address the use of references in a context similar to that found in the job environment. While providing a chart from a handbook is good, giving the trainees handbooks and requiring that the chart be found can be even better. Keep the content of open reference test items valid.
(g) Care must be taken in the development of open reference test items. Under certain circumstances, these may become 'double jeopardy' test items. In other words, if the test developer is testing the ability to use a

graph versus locating it, the trainees will not be able to demonstrate use of the graph if they cannot find it.

An open reference test item is the most challenging kind of test item to develop because it should challenge the highest cognitive level of the trainee. These test items are the result of a great deal of effort on the part of the developer. Several revisions and reviews by others are usually required to refine the test item into a quality test item that will challenge the trainee's decision making ability.

4.5. TEST ITEM ANSWERS AND POINT VALUE

For each test item, a correct answer should be supplied in order to develop an answer key. The answers should be unambiguous so that the grading of the test item will be consistent. The minimum response for full credit must be given and indications of the relative value of partial responses should be made. This kind of information is obvious in multiple-choice and matching test items, but short answer and essay test items do leave room for interpretation.

To prevent interpretive inconsistencies in grading short answer or essay test items, the designated instructor should include the minimum answer required for full credit. If an exact wording or specific concept must be included to fulfil the requirement of a correct answer, one method that may be used is to underline the necessary parts and assign point values to the individual parts. If an exact underlined/italicized wording is not necessary, include a note to the grader that the content is necessary, but that the wording is left to the grader's discretion. For example:

Test item:

State the purpose of the emergency diesel generator. (4 points)

Answer:

Provides *emergency power* (2.0 points) to *essential loads* (1.0 point) in the event of a *total loss of power* (1.0 point). Note: Alternate wording acceptable; grader discretion requested.

After the test item and answer key are written, the point value should be assigned in a clear and consistent manner. The assignment should be based on a point scale with the easier test item assigned lower point value and the more

difficult test item assigned higher point value. The point value should be placed in parentheses at the end of the test item stem.

4.5.1. Cognitive level versus point value

To determine ease or difficulty of a test item, consider the criticality of the enabling objective that is being tested to the overall mastery of the terminal objective which will allow successful performance of the job task. Consider also the cognitive level at which the objective being tested is written. For example, if you are writing multiple-choice test items for three enabling objectives, two of which are written at the knowledge level of Bloom's Taxonomy to support a third which is written at the comprehension level, then the point value assigned to the higher cognitive level would be higher than the point value assigned to the lower cognitive level. For example, this is how these concepts might be applied:

Action verb	**Cognitive level**	**Points**
E.O. 1.1 List	Knowledge	(2.0)
E.O. 1.2 State	Knowledge	(2.0)
E.O. 1.3 Determine	Comprehension	(4.0)

If multiple-choice test items are written for these action verbs, the first two test items would have lower point values than the third. This is because the third is more important, more critical to the mastery of the terminal objective (the first two objectives support the third). Also, the third action verb is at a higher cognitive level (comprehension) than the cognitive level of the first two (knowledge).

Other considerations for determining point value of test items include:

(a) Impact on successful performance of the job task if the answer is not known by the trainee;
(b) Learning level required (lowest level being knowledge and the highest level being evaluation);
(c) Number of answers required for the test item;
(d) Relative degree of difficulty.

Another consideration in the assignment of point value is related to the test item. Test items having multiple parts requiring answers should have at least one point per answer and matching test items should have at least one point per match. Point value allocations should be made and approved when drafted test items are submitted for review and approval.

4.5.2. Scoring essay format test items

A major difficulty in the use of essay format test items is the inability to maintain consistent scoring. The answer must clearly identify the key points of each test item and the grading criteria must be followed closely to ensure consistency of scoring. Effective grading criteria eliminate bias towards trainees (whose answers appear plausible, or who have neat handwriting, or are known to the instructor) by removing subjectivity and promoting objectivity.

During scoring, it is easy to deviate or vary from the established grading criteria. It is, therefore, necessary to check each answer against the answer and grading criteria for every test. It is also advisable for the same person to score one test item across all tests before scoring the next test item. This increases the consistency of the scoring process.

Also during scoring, the examiner may find that a large percentage of trainees incorrectly answer a test item. The assessor should then re-evaluate the test item to determine if it could have been interpreted differently than was intended or if there may have been other potentially correct answers.

4.6. EXAMINATION SPECIFICATIONS

Examination specifications are a blueprint, or plan, that clearly defines the scope and content of the test. It is the documentation for the decisions made in the initial planning stages. Just as it is important to develop training objectives before instruction is planned, so it is necessary to develop test specifications prior to test construction.

The development of examination specifications is a vital step in the testing process. Test specifications provide two important checks on the entire test mechanism:

(1) An explicit, documented link between each test item and a training objective that is verified as being relevant, important and based on the task;
(2) Consistency in the way tests are developed.

Consistency will assist in reducing biases in test content owing to instructor likes and dislikes or the changing of personnel. The process ensures that all decisions regarding job placement are based on trainee performance on the same body of knowledge and ability, even though specific topics covered on individual tests may differ.

4.6.1. Developing examination specifications

Since training objectives complete with action statements, conditions and standards already exist, the major portion of test planning is accomplished. What remains is to determine which objectives will be covered in the test, how many items will be included and which test items are of relative importance. When developing test specifications for examinations, it is important to recognize that the knowledge and skills for all training objectives must be tested at some point in the training.

Table 7 shows an examination specification developed from a list of training objectives. The objective statements indicate the type and level of performance expected of the trainee. The instructor should select the objectives that will be tested on a given examination and establish the relative emphasis each training objective receives.

As Table 7 shows, Objective III.2 is given twice as much weight on the test as III.1 and five times as much weight as III.3. These different weights are based on the objectives' comparable importance to success in job performance and should reflect the relative time spent on the objectives during the course of the training programme. There are no pre-established rules for determining the specific weight assigned to the various areas of test specifications. However, the objectives that represent task elements that are critical to the successful accom-

TABLE 7. EXAMINATION SPECIFICATION

Objectives for training	Testing emphasis (item weight (%))	Objectives to be included in test
I. Area A		
1.	5	Yes
2.	10	Yes
3.	0	No
4.	5	Yes
II. Area B		
1.	10	Yes
2.	0	No
3.	2	Yes
4.	0	No
III. Area C		
1.	5	Yes
2.	10	Yes
3.	2	Yes

plishment of the task must be tested and those test items cannot be missed. The test developer should obtain input from other trainers, from SMEs and from facility operations management and supplement this with his/her own prior experience. Trainees will expect the testing emphasis to be comparable to the emphasis stressed during training and this should be the case. Training objectives can be assigned greater emphasis by increasing the number of test questions for those objectives.

Table 7 further shows that Objectives I.3, II.2 and II.4 do not appear in this test. This may be because they were covered in previous tests, or will be covered in later tests, or can be tested in conjunction with other objectives. All training objectives must be tested at some point during the training (but not necessarily in the final examination) or consideration should be given to their importance to the overall objective of the specific subject area. That is, if it is not considered important enough to be tested, it most likely is not important enough to be included in the training programme.

The completed examination specifications (test outline or sample plan) provide the test developer with a list of training objectives upon which to base test questions. The accompanying CD-ROM contains several administrative procedures for examination specifications.

4.6.2. Examination construction

The actual examination is constructed following test design and test item development. Examination construction requires the test developer to establish the test layout, assemble the test items, prepare the answer key and write test directions.

Test developers should construct tests to some predetermined standardized appearance. The layout and source for this appearance are not as important as maintaining consistency for all of the tests. This consistency of appearance has several advantages. One advantage is minimizing trainee stress by providing a layout trainees are familiar with; test day is an inappropriate time to introduce a new test format or layout. Another advantage is improved reliability. Inherent reliability is based on the consistency or dependability of test results. Trainees should be tested in a similar manner from test to test. This involves similar test duration, consistent grading and scoring, use of the same construction guidelines, consistent coverage of topics, familiar test item formats, etc.

Procedures should establish the format and layout of training programme examinations. Examinations are assembled according to the following general guidelines:

(a) Select the appropriate test items based on the test specifications;
(b) Prepare the test key when the test is constructed;
(c) Indicate and be consistent with point allocations for each answer in regard to the importance of the training objective that the test item is testing;
(d) Assign the number of questions per content area that reflects the appropriate emphasis;
(e) Change the test content from one test to the next so that the tests are not compromised.

4.6.3. Examination layout and assembly

The examination should be assembled in a logical and easily understood format and should follow conventional rules of order for the test items.

Written tests should include typed or printed test items (no handwritten tests) and should be reproduced so that each trainee has a copy of the test. Writing the questions on the board or stating the questions orally invites misunderstanding. An oral examination is not meant to be a written test given orally; rather, it is a unique situation requiring two-way communication.

The test should be clearly labelled. The course, test title, associated unit of study, administration date and test form should be stated on the test. If the test is to have trainee responses written on it, put this identifying information on a cover page where the trainee's name, employee number, or other required information is entered. The preferred arrangement of test items is to group:

(a) All items using a common body of supporting information (e.g. diagram, table, or scenario), even if test item formats must be mixed;
(b) All items of the same format;
(c) All items dealing with the same training objective;
(d) Items from the least to the most difficult.

Some examinations consist of only one format, while others may contain a variety of formats. Although using only one format has the advantage of simplicity and clarity in that it gives only one set of directions, it is more difficult and time consuming for the test developer to force all questions into a single format. There is nothing wrong with a variety of formats; however, to keep the test responses ordered from simple to complex, the following order of test items is suggested:

— Multiple-choice items;
— Matching items;

— Short answer items;
— Essay questions.

When a diagram, drawing, or block of information is used with a test item or items, place it above or below the test question if possible. If it is too large to go on the same page as the test item, it should be attached as the next page in the test so the trainee does not have to search for it. The test item should state the location of the diagram, drawing, etc., if not on the same page. Avoid splitting a test item's material between two pages, but if one is split, present all of the item alternatives on the same page. Keep matching items together on the same page.

Consideration should be given to placing only one question per test page. This minimizes the administrative burden on the trainee, improves test clarity and reduces the chances of the trainee inadvertently failing to answer a question.

4.7. EXAMINATION BANKS

Training organizations typically develop and maintain examination banks. These banks consist of previously used tests, answer keys and test items. Not only do these test banks save a great deal of time, but the resulting tests are significantly improved because of any modifications made following the use of each test. Training programmes should include such a test bank and instructors should collect test analysis information each time a test is conducted. Since training organizations may provide training by programme area using several instructors, it is important that the test bank concept be applied at the programme level. In this way, the size, scope and uniformity of the testing process will be improved.

The widespread use of computers and database software has added significantly to the capabilities and flexibility of such systems. For example, multiple versions of a test may be produced to increase test security during administration. There is a large amount of written test generation and records maintenance software systems available to increase the ease and efficiency of test development and administration. These systems also provide an effective tool for test item analysis and improvement.

The following should be considered when establishing test banks:

(a) The scope of the bank.
(b) Effective security controls for computerized test banks.
(c) An ongoing programme for test and test item analysis.

(d) The use of machine scored answer sheets as appropriate.
(e) Clear guidelines and procedures for use of the test items.
(f) A test outline or test specifications.
(g) A test item numbering system. The following is a list of potential test item identifiers:
    — Programme;
    — Procedure number;
    — Lesson plan;
    — Training objective;
    — Test item format;
    — Test item level of difficulty;
    — Point value;
    — Date test item is generated;
    — Test item generated;
    — Dates test items are used on tests.
(h) An ongoing programme for test item review, replacement and introduction of new test items.
(i) Sharing information with other training organizations.

The accompanying CD-ROM contains several examples of examination banks.

4.7.1. Trainee access to examination banks

Examination banks can be a valuable source for learning as well as testing. However, if the number of questions in the examination bank is too small, trainees could simply study former examination questions. The trainees' performance in an examination could be compromised because the cognitive level at which the trainees are tested would be decreased to the simple recall level. Comprehension and analysis levels of knowledge would not be measurable because the mental thought process has been reduced to simple recognition. Because of this concern, many training organizations have placed limits on the availability of examination bank questions to trainees. In order for the examination banks to be available to trainees it is usually required to have a minimum number of test items in the examination bank. Typically, 700 or more questions should be in an examination bank before trainees are allowed access to it. In addition, when examinations are prepared, no more than 75% of the questions are to be taken directly from the examination bank. The remaining test items are then either new test items or revisions to ones. In addition, it is also usually required to maintain a dynamic examination bank by

requiring the review, revision, or generation of a minimum number of questions (e.g. 150) each year.

4.8. WRITTEN EXAMINATION ADMINISTRATION

Examination administration has an important effect on the usefulness of test results and requires control. The instructor should ensure that a suitable environment is established, consistent and clear test directions are given and proper supervision is present for the entire test.

4.8.1. Establish environment

Effective testing environments require attention to the physical qualities of the test setting and to the trainees' emotional 'climate'. High noise levels, poor lighting, lack of ventilation, excessive heat or cold and frequent interruptions will lower trainee test performance. The assessor should optimize, to the extent possible, the conditions for testing. This may be as simple as scheduling testing in the morning if the classroom becomes too hot in the afternoon.

While most assessors are aware of the physical testing environment, many do not give sufficient consideration to the emotional environment they establish. The testing environment should be conducive to effecting testing. Making the purpose of the test clear and emphasizing the need for accurate test results can create a good emotional climate, which is important in building motivation, reducing anxiety and improving communications.

4.8.2. Test directions

Each test should have clearly written directions. These directions should tell the trainee what to do, how to do it and how to record the responses. General directions should be given for the test, with specific directions given for each section, subpart and item format. Though the assessor should present the directions orally prior to the start of the test, the written directions should be clear enough to enable the trainees to complete the test without any further instruction. The trainees should be given time to read the instructions and ask questions before the test is started.

Questions that require mathematical calculations pose a unique problem. Suppose a trainee performs a complex equation using a calculator in one step, while the answer key breaks down the calculation into individual steps. The resulting answer will be different from the answer provided in the answer key, since the answer key will break the answer down into individual steps. Each

step is then calculated separately and rounded to a significant digit. Rounding of answers (or individual step answers) can cause an otherwise correct answer to be marked as wrong, because the answer key specifies a discrete number. Therefore, precision or accuracy of answers needs to be addressed in the test directions and in the answer key.

Inform the trainees that they may ask questions during the test. Avoid giving individualized assistance by providing any clarifying information from individually asked questions to the entire group.

Trainees should be told the value of test items and how they will be scored. The trainee should know whether partial credit will be given, what degree of precision is required, whether units must be identified (such as bar, ohm, rem) and, in the case of calculations, if work must be shown. Time limits should be stated.

When developing the instructions, keep them clear and concise. Make important points stand out by using a different size type, placing the type in bold, or by underlining. Have an independent review done of the directions to check for inconsistencies or potential misunderstandings. Consider including sample items with the directions when introducing difficult or unusual item formats. Clear directions will help maintain the reliability and validity of the test. Appendix VI provides an example of test directions to trainees.

4.8.3. Test monitoring

Effective test monitoring will ensure that everyone has the same opportunity to understand and answer the questions properly. It is important that the test results provide an accurate indication of a trainee's performance.

Training procedures should provide definitive guidance for test monitoring. A clear policy on academic honesty should be established at the beginning of any training programme and should be enforced throughout the programme. The single best method is to observe trainees carefully during testing. Some training department procedures require that each trainee sign an affidavit, usually on the test cover sheet, stating that the work is the individual's own. This has some deterrent value; however, it should not be allowed to replace other useful methods. These include spacing trainees during testing, using multiple test forms and revising the test for each session.

4.8.4. Use of handouts

The way in which handouts are incorporated into the examination has a bearing on the reliability of the examination. If a trainee has to search for the applicable drawing or curve, it adds to their frustration level and the time

consumed in test administration. The format of the examination itself also contributes to its reliability. Not only does it take more time to complete an examination that incorporates a multitude of administrative requirements, but it also detracts from the trainee's ability to focus on just the examination test items.

Include handouts such as matrices or incomplete system drawings on which the trainee must transcribe his/her answer directly behind the page of the applicable test item. This will help ensure the integrity of the trainee's package of answers.

Include all other handouts, such as reference material, operating procedures, technical specifications, excerpts from technical manuals, etc., within a separate handout package. This package should be page checked with the trainees before examination administration. Booklets, rather than copied pages, should be used whenever possible.

Use of the single page/single test item layout should be considered. This has proven to facilitate both the examination grading and the test taking processes. It also minimizes the administrative burden on the trainee, greatly reducing the chances of any test items being missed.

5. ORAL EXAMINATIONS

Oral assessments in some form are an integral part of any qualification process. Knowledge of a subject (theory, operational interrelationship, etc.) is always required for qualification and oral examinations are an effective method of determining whether the person understands the job, the task, or the evolution.

Although this section focuses on purely oral examinations, oral questioning techniques also form an important element of performance based assessments (Section 6), such as questioning the trainee during the operation of a plant system to check their knowledge of procedure bases, expected plant response, expected indications, etc. Similarly, questioning techniques may be used as part of simulator assessments to check underpinning knowledge, understanding of system interactions, etc. Oral assessments may also be used to supplement written examinations, e.g. to check unsatisfactory/ambiguous answers or to clarify remedial training requirements.

Oral examinations or oral 'checkouts' (checkout is a term that is used as a synonym for the word examination (see the definition in Appendix I)) are used

in all training and qualification programmes. For example, the simplest form of oral examination involves the questions that are asked of a student, trainee, or incumbent as the individual demonstrates how to perform a simple task. This is typically the knowledge assessment phase of OJT.

Another example of oral examination is a formally convened and administered board consisting of a group of experts asking questions to determine the person's knowledge relative to the position for which qualification is sought. Other forms of oral examination are also used. These include oral checkouts, facility walkthroughs, operational examinations, performance demonstrations and others. If oral examinations are properly developed and consistently administered, they are useful to test the level of learning and comprehension.

5.1. BENEFITS OF ORAL ASSESSMENT

The main benefit of oral assessment is that it allows the assessor to probe the depth of knowledge and understanding of a trainee in a way that is difficult to achieve with written assessments. It also allows the assessor to adapt the questions dynamically according to the trainee's response, although this must be within the bounds of the training objectives.

5.2. TYPES AND USES OF ORAL EXAMINATIONS

Several types of oral examination are used for the qualification or certification of nuclear industry personnel. The most common include the oral checkout, oral board and performance demonstration or operational evaluation (the latter is discussed in Section 6). Regardless of the type of oral examination used, questions are based on training objectives derived from an analysis of the job or of the training course requirements.

5.2.1. One-on-one oral checkout

The oral checkout is an oral examination usually conducted on a one-on-one basis (assessor to trainee). Oral checkouts ascertain an individual's fundamental knowledge of theoretical concepts, equipment and system operation, normal and abnormal operating procedures and interrelated systems or components. Oral checkouts should use the training objectives as the basis for the questions.

Appendix VII is an example of a checklist for a maintenance task that includes both a performance assessment and one-on-one oral questioning. Appendix VIII is another example of a specification for a workplace assessment that includes both an oral and a performance assessment. Several examples of the procedures for developing OJT checklists are included in the accompanying CD-ROM.

There may be special cases in which a written examination is administered using an oral checkout. In these cases, a written examination is given to a trainee by reading the questions to the trainee. The responses should be recorded for the trainee by the examination proctor or verified against the correct answer. When the trainee is required to read and understand facility procedures and warning signs to perform the job, oral administration of written examinations in related areas should be prohibited.

5.2.2. Oral (or examination) board

The oral board is an oral examination where a panel is convened to determine if a trainee has achieved the level of knowledge required for qualification. An oral board may be used as the final check of qualification, or it may be combined with an operational assessment or plant walkthrough examination. Oral boards used in a nuclear facility usually cover all facets of the facility and process operations. This may include (but is not limited to):

(a) Facility components (including design base, functions, etc.).
(b) System interrelationships.
(c) Normal/abnormal situations involving systems and interrelated systems and components.
(d) 'Soft skills' aspects, such as communications, teamwork, attitudes, etc.

Oral boards can be a good opportunity to clarify any outstanding issues from written and/or performance based examinations. Some of the questions asked may be based on ambiguous answers from these earlier examinations.

5.2.3. Other uses of oral boards

Two other types of oral board are sometimes used and are commonly referred to as progress review and programme evaluation boards.

5.2.3.1. Progress review board

The progress review board can be used to assess a trainee's progress towards the completion of assigned qualifications or to determine if a trainee possesses the knowledge to enter a higher level of qualification. Progress review boards may also be used to determine whether or not a trainee is having difficulty with the training programme or has a reasonable probability of completing the programme. Problems with meeting the minimum progress requirements may be due to personality conflicts or to a lack of understanding of fundamental principles and concepts. The progress review board can be used to identify the cause of the problem. It may also uncover problems with the training programme that had been missed during programme evaluation. This discovery could lead to using the programme evaluation board to examine most or all of the trainees in order to determine the extent of the problem. The progress review board is more valuable early on in the qualification process rather than near the end. It also serves as training in board techniques for both trainees and prospective board members. Progress review boards should follow the guidelines for administering oral boards as discussed in the Section 5.6.4.

5.2.3.2. Programme evaluation board

The programme evaluation board can be used for identifying problems in the training programme. When a problem is identified with a training programme (e.g. incidents involving qualified personnel, slow qualification times compared with the past norms, abnormally high examination failure rates), a programme evaluation board could be used to identify the cause of the problem. The programme evaluation board should also provide recommendations for corrective action (e.g. re-teaching the required material) that are based on the causal factors. Programme evaluation boards need not follow the guidelines for administering oral boards. For example, the programme evaluation board may be administered to several trainees at the same time and board membership should consist of personnel who are responsible for the training programme in question. Typically, membership should include line organization and training organization personnel. Strictly speaking, this is not an oral assessment of trainees, although trainees are questioned, but rather it is an evaluation of training content, based on the questioning of training participants and therefore should be considered in conjunction with Sections 6 and 8.

5.3. GENERAL GUIDANCE FOR ORAL ASSESSMENTS

5.3.1. Procedures for oral examination

Procedures should be developed that describe how to develop and administer all types of oral examination used. These procedures should address the following issues:

(a) Materials that will be available to the trainee during the examination (these should include materials that are normally available to a person during job performance);
(b) Qualification of personnel who conduct each type of oral examination, or how a person is approved to conduct oral examinations (this includes determining the make-up of an oral board);
(c) Management of non-typical situations (e.g. interruptions of oral examinations due to unexpected facility conditions, high background noise);
(d) Methods of grading oral examinations to ensure reliable, consistent and equitable oral testing;
(e) Methods of assigning relative importance to questions (e.g. questions relating to significant facility processes and safety systems are valued higher than questions relating to non-vital auxiliary systems or components).

5.3.2. Examination intent

Oral examinations should probe the trainee's understanding of fundamental principles, the ability to apply these principles to practical situations, knowledge of equipment and system operation and normal and abnormal operating procedures. Examinations should be based on knowledge of information within the scope of the training objectives.

An effective way to probe understanding is to ask the 'whys', 'hows' and 'what ifs'. The following are examples of the types of question that can determine a person's understanding:

(a) Why is the equipment built as it is?
(b) Why does step one come before step two and what would be the consequences of reversing the order?
(c) Why does the process work as it does?
(d) How does the facility respond to specific action?
(e) How does the operator respond to a specific facility condition?

(f) What happens to process indications in system A if system B does this (describe the action)?
(g) What action(s) can the operator take if process indications show this symptom (describe the symptom)?

5.4. DEVELOPMENT OF ORAL QUESTIONS

Questions should be asked which test the trainee's knowledge at all appropriate cognitive levels. Questions involving a choice between alternatives, or involving situations not directly covered by a procedure, are examples of questions testing the higher cognitive levels. When answering questions concerning abnormal events, a trainee's answer should be consistent with the requirements set by facility procedures. Situation problems should be described as they actually occur in the facility.

When probing the trainee's level of knowledge of the facility operations area, the examiner should ask the trainee questions relating to both normal and abnormal situations. For example, to see if a trainee really understands what happens during a loss of control air, the examiner asks:

(a) What are the loads on the control air system?
(b) What happens if normal control air is lost?
(c) What happens if the backup compressor does not work properly?
(d) What happens if the startup procedure is not followed?
(e) What should an operator do if a normal corrective action fails to rectify a problem?

A line of questioning such as this gives the trainee ample opportunity to use collective knowledge and explain how things work.

Trick (deceptive, misleading) questions with hidden meanings should never be used. The assessor who repetitively questions many trainees on the same familiar material may tend to think that the fundamental, straightforward approach is too simple. This may lead the assessors to develop clever or unusual (trick) questions. A trainee exposed to this type of questioning technique may be inclined to search for hidden meanings in subsequent straightforward questions. This tendency may adversely affect the trainee's ability to express a true level of understanding.

Questions and answers for the oral examination portion of OJT assessments should be prepared in advance during the development of the job performance measure (JPM) or other instrument used to assess OJT. The primary and planned follow-up questions and answers should be prepared in

advance. Any follow-up questions, generated by something the trainee said, should be written down along with the answer on the JPM at the time they are asked. Any follow-up question should be tied directly to their primary question.

5.4.1. Types of question

The examiner has the option of asking several different types of question during an oral examination. The two most common types of question are open-ended and closed-ended. The open-ended question places the burden of conversation on the trainee. The trainee brings all known information together to answer the question. Examples of open-ended questions are:

Example 1. What are the reasons for starting the recirculation pump with the discharge valve closed rather than open?

Example 2. Why does the safety system actuate at that pressure?

Closed-ended questions are used to elicit specific answers. They should be worded so that they are not answered as yes or no, but with an answer that demonstrates knowledge of the component, system, procedure, etc. An example of the yes/no question is: Does the pump have a white 'POWER ON' light? Example 3 is the preferred method of asking this question.

Example 3. What indications does the main control panel operator receive that the recirculation pump motor controller is energized?

Another example of a closed-ended question is:

Example 4. What indications of a loss of pump prime are available to you at this control panel?

The examiner can use closed-ended questions to clarify a statement that the trainee makes in response to an open-ended question. Using a mix of question types should provide the assessor with enough information to determine whether the trainee has adequate knowledge.

5.4.2. Questioning process

Questions should not be asked in a manner which tends to lead the trainee to the answer. One of the reasons for having the original questions and

any planned follow-up questions written prior to the examination is to help prevent asking leading questions. The assessors should state the question clearly, as written, and then give the trainee reasonable time to think and answer. Only one question at a time should be asked. Additional or follow-up questions should not be asked before the trainee has had time to answer the original question. Additional questions could distract the trainee or could lead the trainee to the answer.

The assessor(s) should encourage the trainee to explain what the trainee knows. For example, ask the trainee to apply a principle of good use of facility procedures to determine the proper action in a specific practical situation. This ensures understanding of the principle rather than just memorization of the words of the concept.

If a trainee gives an incorrect, partially correct, or unclear answer to a question, the topic should be probed further to establish the trainee's true level of knowledge and understanding. Questions should be clarified and restated as needed in response to trainee requests. The assessor should ensure that restating a question does not change the original question's intent or training objective basis and does not lead the trainee to the correct response.

5.5. ADMINISTRATION OF ORAL CHECKOUTS

Assessors should utilize the associated training objective(s) when conducting oral checkouts. Generally, the training objectives will specify that the trainee demonstrate knowledge using the stated reference. Assessors should ensure that the trainee has access to, and uses, references such as system diagrams or procedures during the oral checkout.

Oral checkouts should include an in-depth questioning of the trainee's understanding of fundamental theory, system equipment and processes and a discussion of procedures as required by the learning objective(s). Special emphasis should be placed on the trainee's ability to apply this knowledge to facility operations. Questions asked during oral checkouts should not be restricted to verbatim restatements of the training objective(s), but may be rephrased. The oral checkout may be used to test understanding and judgement, as well as factual knowledge.

The length of time taken for an oral checkout should depend on the requirements of the training objective(s). Oral checkouts typically last from 5 to 20 minutes, depending on the assessor's and trainee's preparation and the complexity of the subject. Facilities should not set specific time limits for oral checkouts; the checkout should last as long as necessary for the assessor to determine that the level of knowledge of the trainee is consistent with the

requirements of the training objective. However, an oral checkout should normally be completed in one session.

5.5.1. Assessor actions at completion of an oral checkout

At the completion of an oral checkout, the assessor should make an objective determination, referring to the training objective(s), as to whether the trainee has a satisfactory level of knowledge. The signature of the assessor indicates that the trainee has successfully completed the oral checkout and that the trainee possesses the minimum required knowledge. Several possible outcomes from an oral checkout are:

(a) Satisfactory with no deficiencies: The trainee displayed a satisfactory knowledge showing no weak points of any significance and receives a signature.
(b) Satisfactory with minor deficiencies (corrected to satisfactory): The trainee demonstrated a good overall understanding of the topic, but lacked information on minor details. The assessor should teach the trainee those details and consider the checkout to be satisfactory or have the trainee review those details and return with the correct answer to those questions. The same approach should be used if the trainee has a lack of understanding on some principles for which a reference is not readily available. After the details are discussed to the satisfaction of the assessor, the trainee receives a signature.
(c) Unsatisfactory: The trainee has a significant lack of understanding of the material. In this case, the trainee should not receive a signature and should be counselled as to (1) what material to restudy and (2) the best method of studying it (including scheduled retraining if necessary) before receiving another checkout. The assessor should establish the date for the trainee to return for the checkout. Establishing a date for the checkout gives the trainee direction and a goal for the completion of the restudying.

5.6. ADMINSTRATION OF ORAL BOARDS

5.6.1. Oral board preparation

The trainee's manager and/or the training manager (or other persons as described in the training programme description) should determine when a trainee approaches the point in their qualification programme where an oral board is deemed appropriate. Management selects the oral board members and

chairperson, as specified by facility procedures. The location of the oral board should be selected using the guidelines outlined in the Section 5.6.3. The board date should be established well in advance so that the trainee and board members have sufficient time to prepare. The board chairperson should also schedule time after the board with the training manager to discuss any significant problems identified in the trainee's knowledge.

Before convening an oral board, the board chairperson should ensure that board members meet the requirements to participate on the board (see Section 5.6.2). The training department and board members should review the trainee's training records to verify that all prerequisites, such as OJT, oral checkouts, walkthroughs and written examinations, have been completed and properly documented. This prevents examination of a trainee who has not completed all the requirements for qualification.

Oral board panels typically consist of three to five members. Individual members ask questions to probe and assess knowledge. To prepare for the board, each board member should prepare questions (with answers) to be asked during the board. Questions should represent a cross-section of the material contained in the training objectives. Follow-up questions (with answers) should also be prepared prior to the board.

It should be noted that the guideline of having prepared questions and follow-up questions should not prevent the board member from developing and asking probing questions (with answers) during the conduct of the board. These questions, however, are to be tied directly to the topic covered by the prepared questions.

The questions and answers may be selected from an examination (test item) bank or developed from the training objectives. The questions that are not taken directly from an examination bank should be reviewed by the training organization for relevancy to the training objectives before the board convenes.

Board members should bring copies of their prepared questions and answers to share with the other board members. This practice helps prevent grading differences between board members.

The accompanying CD-ROM contains examples of administrative procedures for the development and administration of written, oral and simulator examinations.

5.6.2. Participation on oral boards

Oral examination procedures should include a section that specifies who can participate on oral boards and who can act as the board chairperson. This section of the procedure should also specify the voting process and should

accommodate both the presence and absence of participants from other organizations, such as the regulatory body. There should be a separate oral board membership list for each job position. Persons participating in the board should be at an equal or greater qualification level than the trainee. A person with appropriate training expertise should always be present to ensure consistent application of the training and qualification requirements and that questions properly relate to the training objectives being tested. Participants from external organizations who function as co-assessors should follow the same guidelines as other board members with respect to preparation, participation and grading. They should also only grade in areas of their expertise.

The number of persons present during an oral board should be limited so as to ensure board integrity and to minimize distractions to the trainee. Other trainees should not be allowed to be present during an oral board. Other assessors may be present either to witness the oral board as part of their training or to audit the performance of the assessor(s) administering the oral board. Others may be allowed to observe the oral board if (a) the board chairperson approves their presence and (b) the trainee does not object to their presence.

The conduct of observers should be described in the facility procedure on oral examinations. Observers should neither enter into the questioning, nor interrupt the proceedings. Any comments they wish to make or questions they have concerning the conduct of the board should be held over until the examination is completed and the trainee is out of the room or area.

5.6.3. Oral board facilities

Facilities should have a room that can be used to conduct oral boards. The room should be designed in such a way that it provides an atmosphere that is conducive to conducting an oral board effectively and without interruption from outside sources. Items to consider when establishing an oral board room include:

(a) Black/green board and chalk, or a white board and markers are available for use by the trainee.
(b) All reference materials that are normally available to the trainee, when qualified and on the job site, should be placed in the board room prior to the examination.
(c) ‘Beepers’ or pagers should be turned off during the oral board, if possible.
(d) Telephones should be disconnected.
(e) Interruptions should be sanctioned on an ‘emergency only’ basis.

5.6.4. Conduct of oral boards

The board chairperson ensures that all required training or operations documents necessary for the board are present. When the oral board chair is prepared to begin, the trainee is asked to enter the board room. The oral board chair should cover the board guidelines with the trainee. Appendix IX, Example Oral Board Instructions, is an example of oral board instructions provided to the trainee.

Oral board members should actively participate in questioning the trainee. The questions may be asked about one area at a time, or in a round table fashion with board members asking questions in their areas in turn. However, each board member should grade only subject areas in which he/she is qualified (as determined by facility procedure) to grade. Board members should refrain from coaching the trainee to an answer by head nods, smiles, head shaking or jests.

The board chairperson should ensure that the board is conducted in a formal fashion and that all board members conduct themselves accordingly at all times. The facility procedure on oral examinations should be present during all oral boards. The board chairperson should ensure that no discussions concerning answers, grades, materials or opinions take place until the trainee leaves the room. Board members should not discuss grades until all board members have completed grading the trainee. Each board member's grade should represent an independent assessment of the trainee's performance (assigning grades in the topic areas and overall board grade are discussed in the next section).

An oral board may last from one to three hours, depending on the complexity of the qualification and the level of qualification (non-supervisory boards being shorter and supervisory boards being longer). If the board lasts too long, the trainee and oral board members will not be able to function efficiently. In all cases, questioning should continue until all board members are satisfied that they have determined the trainee's level of knowledge and can accurately grade the trainee's performance. At the conclusion of the board, the board chairperson should ensure that a representative sample of the training objectives has been covered in sufficient detail to allow determination of the trainee's level of knowledge.

Breaks can be allowed during oral boards. Subjects covered before a break should not be discussed after the break. At least one board member should remain with the grading and question/answer sheets at all times to prevent examination compromise.

At the conclusion of questioning, the board chairperson should dismiss the trainee. The oral board members should then tabulate their individual

grades and determine the trainee's pass/fail status. When the associated paperwork is completed, the board chairperson recalls the trainee and informs the trainee of the board's decision. The board chairperson should discuss highlights of the results of the board with the trainee (discuss all areas in which the trainee may need to study to correct any identified knowledge deficiencies and indicate areas of excellent performance). The trainee should be encouraged to ask questions regarding the outcome of the board.

5.6.5. Oral board grading

The grading standard for pass/fail should be set by facility procedure. Each board member should independently grade each question that the board member is qualified to grade. The grades should be recorded on a standard form by each board member and the overall board grade assigned by the board chairperson. These forms, completed and signed by each board member, should become a part of the trainee's training record.

Board questioning should be organized into distinct areas such as theory, systems and components, normal operations, abnormal operations, administrative controls, technical safety requirements, etc. Questions pertaining to each area may be asked by one or more of the board members. Assigning a grade to each question, to each area of questioning and to the overall board can be difficult. Appendix X contains a detailed example of an oral board grading methodology.

If remediation training is assigned by the board to correct noted deficiencies (remedial training may be given whether the trainee passes or fails the board), board members should review their individual question record sheets and assist the oral board chairperson in determining the required remedial actions. Remedial actions may vary from minor points discussed with the trainee directly after the board to more extensive requirements (formal remediation that is tracked and evaluated). The specific remedial action plans should be developed by the trainee's training organization and approved by line management. The training organization should document and track all remedial actions assigned to the trainee.

6. PERFORMANCE ASSESSMENTS

Performance assessments or tests measure task performance in the job environment and serve as a mechanism for determining task qualification.

A performance test consistently and systematically evaluates the ability of the trainee to perform a task. Asking trainees to describe proper welding techniques is not a performance test; asking trainees to make a proper weld is. The performance test is not a training instrument; rather, it is a testing tool that ensures consistent performance evaluations. However, some Member States provide trainees with a 'student copy' of the performance test for use as a training or study aid. The student copy contains all the same information included in the test instrument except the answers to the oral questions used for assessing trainee knowledge. A performance test should test both the knowledge and practical requirements that were derived during the analysis of the task.

Performance assessments are most successful when carried out in the actual work situation. Where this is not possible (e.g. for reasons of safety, access, downtime), a simulated workplace may be used. Simulated workplaces are designed to replicate, to the extent possible, the actual workplace (see Section 6.1.4, Determine Method of Accomplishment).

Persons who administer performance tests are typically required to be trained and qualified as performance test assessors. The qualification of assessors ensures consistency of performance examinations and application of management expectations throughout the organization.

JPMs or other OJT assessment instruments are typically methods used for the development of objective based training materials. The JPM identifies the elements (procedural steps), knowledge and skills necessary to perform the task. It also identifies the initiating cue that prompts or signals the trainee to begin the task, the terminal and enabling objectives and the conditions under which actions occur and establishes standards that measure satisfactory performance of the elements and thus the task. Typically for authorized control room operating positions, the JPMs or OJT assessment instruments are fairly detailed. While for other job positions, the OJT instruments are much simpler and rely to a greater extent on the training and experience of qualified OJT trainers and assessors. Reference [7] contains examples of JPMs that are used for authorized job positions. Section 6.3 addresses other methods of both oral and performance assessments and includes examples.

6.1. DEVELOPING PERFORMANCE ASSESSMENTS

Developing performance assessments involves identifying equipment availability and other limitations, determining the best instructional and test methods and constructing a test that provides the most effective measurement of the task. The task statement identifies the task to be evaluated by the

performance test. The OJT or JPM assessment form includes the elements of the task and other supporting information needed for competent performance of each element of the task. The references identified in the OJT/JPM form should be available to the developer when writing performance tests. The developer may choose to make provisions for references, tools and equipment that are supplied at the time of the trainee test or direct the trainee to gather these resources as part of starting the test.

The following steps should be performed when developing a performance test. The test development process should:

— Determine the test limitations;
— Determine the elements to be tested;
— Determine the conditions and the standards;
— Determine the method of accomplishment;
— Construct the performance test;
— Determine the scoring procedures;
— Pilot the performance test;
— Approve the performance test.

6.1.1. Determine test limitations

The first step in developing a performance test is to review the task and determine the potential test limitations. Test limitations are those factors that can have an impact on the development or the conduct of a performance test. These may include availability of time, workforce, equipment and resources. If performance of a task would require more time than is reasonable, then the performance test should be developed using only the critical task elements. Workforce availability can also impose limitations on task performance. These constraints occur when more than one individual is required for task performance.

Situations occur when equipment or facilities will not be available to support the test. Cost can also affect performance tests. The cost of performance test administration and its effect on consumable repair parts should be kept within reasonable limits. Many infrequently performed tasks cannot be performed for training or testing purposes in the job environment. Safety is another factor to consider. If the testing of certain tasks would impose unreasonable demands on the personnel, facility or equipment, test those tasks using simulation as the method of accomplishment.

The items to be considered when determining the need for simulation or some other method of accomplishment in performance testing are summarized as follows:

(a) Downtime: Effect of task performance on equipment readiness and efficiency.
(b) Damage potential: Damage to plant equipment and personnel.
(c) Cost: Cost of using plant personnel, equipment and materials.

If any limitations result in a change of content in an established test, managerial, administrative and instructional approval should be required for the change.

6.1.2. Determine elements to be tested

The elements of the task represent an important design consideration. The developer should determine which elements can be tested realistically and should focus on elements that have the greatest number of skill and knowledge statements. When limitations make undertaking the task during the performance test unrealistic, task elements should be examined. Elements that include important decision points are predictors of successful performance of the task. If they can be tested realistically, they should be included in the performance test.

A critical task element is defined as any element of the task that is deemed crucial to the satisfactory performance of the task. Task elements such as removing interference, obtaining the procedure and cleaning the job site are typically non-critical task elements. Non-critical elements are generally administrative controls and tend to be generic to other tasks. The critical/non-critical designation becomes important in the scoring and assessment criteria. To determine if an element of a task is critical, consider the following guidelines. An element may be critical if its omission or improper execution:

(a) Causes or could cause damage to any system or component to the extent that it prevents the system or component from being immediately available for its intended purpose;
(b) Causes or could cause a serious injury or pose a hazard;
(c) Results in incomplete task performance;
(d) Violates security;
(e) Results in an out-of-tolerance condition or measurement which prevents the equipment from meeting facility procedures or specifications;
(f) Violates a standard maintenance procedure such as improper use of test equipment or hand tools (this does not include performing procedure steps out of sequence);
(g) Causes excessive delays attributable to insufficient job knowledge or improper planning although the task was successfully performed;

(h) Results in delay(s) due to unnecessary troubleshooting, removal or replacement of components or rejection of serviceable equipment.

There are some task steps that must be performed in the proper sequence. These should be identified on the checklist for the instructor and the trainee. These steps can be marked with an 'S' to indicate that they must be performed in sequence.

6.1.3. Determine conditions and standards

After test limitations and element designations have been determined, identify the conditions and standards needed for task performance. Ideally, the test developer should duplicate the cues, conditions and standards of the actual task, although some compromise may be necessary. For tasks with multiple conditions and branching decisions, multiple performance tests may have to be developed.

Conditions are prerequisite requirements that exist for successful task performance. Conditions define facility conditions and include information and resources available to the trainee during task performance. If limitations prevent using all conditions identified in the JPM or OJT assessment form, a sample should be used that best assesses the ability of the trainee to perform the task under actual conditions. Task conditions may require modification if the task cannot be performed under actual conditions. For instance, conditions could include high radiation areas and other environmental concerns.

Performance tests include standards of measurement that are applied consistently in the evaluation of task performance. Standards may relate to the process, the product of performance, or a combination of both. Process standards are step-by-step procedures that must be followed, usually without deviation. Product standards prescribe output (the product of performance) and criteria for judging acceptability of the performance (i.e. surface machined to a tolerance of ± 0.05 mm).

Task standards should be identified in the assessment form or preferably by the facility procedures whenever possible. However, limitations in the test environment may require a best approximation of the job standard used during the performance test. Typically, the conditions and standards for the elements of a task are implied in the conditions and standards of the entire task. However, if an element has a unique condition and/or standard that is not implied, then it should be stated with that element.

6.1.4. Determine method of accomplishment

Each task that is tested should have a designated method of accomplishment, or level of performance, which dictates how the trainee is to demonstrate the task to the instructor. The method is identified for a task and should be identified for the individual task elements so that each trainee is tested in the same manner. There are four possible methods of accomplishment:

(1) Perform the specified task using approved procedures and observing all applicable safety and administrative requirements. This includes a thorough discussion (usually prior to performing the task) addressing safety implications, elements involved, the effects on associated equipment or systems and abnormal situations which may arise while performing the task. This method of accomplishment is the most desirable level for performance testing.
(2) Simulate performance of the specific task. Using approved procedures, 'walk through' the task and simulate all actual manipulations (valves, switches, tools, etc.) an employee would perform. Describe applicable safety and administrative requirements and the parameters (meter readings, charts, measurements, etc.) an employee would observe/monitor during actual performance of the task. Conduct the same discussion as required for a perform task.
(3) Observe an individual performing the specified task. Conduct the same discussion as that required for a perform task.
(4) Discuss the specified task using applicable procedures, piping and instrumentation drawings, blueprints, etc., including the discussion as required for a perform task. Demonstrate knowledge of the task by describing the manipulations required and the parameters that may be expected to change. This method of accomplishment is the least desirable for performance testing.

Simulate, observe and discuss should be used only when actual performance is not feasible, such as in a high radiation area. To demonstrate success with the task the trainee should always perform each of the critical task elements by the designated method. For example, if the task method of accomplishment is perform, then the trainee must actually perform each element designated as critical. The trainee can neither simulate nor discuss those items. The non-critical elements could have a discuss method of accomplishment designation to save test time and allow concentration on the critical items.

Non-critical elements need not be included if focus is required on the critical elements in order to save time.

6.1.5. Construct the performance test

The performance test can be constructed on the basis of the previous information. A performance test typically consists of major items which include:

(a) A performance training objective (task statement) indicating the action and the object;
(b) Condition(s) under which the action is to be accomplished;
(c) Standard(s) against which performance is to be measured;
(d) References;
(e) Method of accomplishment (perform, simulate, observe, or discuss);
(f) Elements (at least critical elements and non-critical if desired) to be accomplished and references indicated;
(g) Knowledge requirements which consist of the cognitive items supportive of the practical requirements;
(h) Practical requirements which consist of the task elements and their related standards.

The questions used for the knowledge requirements should be placed within the assessment standard to indicate when they are to be asked. Directions should require the instructor to read the questions exactly as written. Space should also be provided to record the trainee's response (if the correct response is not given). The correct answer should always be included with the question.

Additional information from the assessment form may be included in the performance test, such as identifying the task's work group and other information as appropriate. Appendix XI, Performance Test Construction Checklist, is an example of a checklist that can be used when constructing a performance test.

The performance test should rely on the use of actual plant procedures whenever possible. When necessary it should summarize the procedure and be designed to evaluate critical aspects of a particular task. If a task requires specific values such as torque and tolerance, they should be stated in the standard for the task or element, if not already included in the plant procedure. Hold points should be inserted at desired locations in the performance test to allow the instructor to grade the trainee's performance of the previous steps.

The performance test package should consist of an administrative section, instructions to both the instructor and the trainee, a guide for the instructor to use for scoring, a trainee guide, an optional data sheet for trainee use and a section used for documentation (e.g. a checklist).

Appendices VII–XIV are examples of various performance tests. These examples vary in the level of detail and content depending on differences in job positions, task importance and difficulty and the methodology used by various Member States.

6.1.6. Develop scoring procedures

The developer should create an evaluation instrument that the instructor can use to measure accurately the trainee's performance of each step of the performance test. The evaluation instrument must measure the trainee's ability to demonstrate the task.

When the performance test is constructed, scoring procedures are developed. A detailed, step-by-step description of required performance provides an effective scoring procedure for some tasks. Action steps or elements required in the performance test are usually prepared in checklist form and the trainee is required to follow each step without deviation. For other tasks, the product of performance (i.e. a tangible result) should be measured. In developing a scoring procedure for this type of performance test, scoring characteristics must be defined to distinguish clearly between satisfactory and unsatisfactory performance.

Scoring methods must adhere to administrative and instructional guidelines and reflect the assessment standards. If an assessment standard is without error, a yes/no checklist should be used. If the standard implies some range of acceptable performance, a rating scale may be used. However, rating scales introduce greater subjectivity and are more difficult to use, to interpret and to back up than a yes/no checklist. If sequence is important, identify this on the performance test and provide proper scoring guidance.

Establish cut-off scores to meet the performance standards. Percentages are the least preferred method. Failure of a performance test should be determined by the failure of any critical step or the failure to follow required sequences. The cut-off score for any behaviour should be based on a level of performance that is achievable by the average person.

All methods of scoring should be consistent with policy, procedures, specifications and needs. The method should also adhere to instructional guidelines, such as testing the objectives, and should clearly distinguish between satisfactory and unsatisfactory performance.

6.1.7. Piloting the performance test

The purpose of piloting a performance test is to ensure that the test can be conducted as planned. The pilot should be conducted under the conditions required for actual job performance or under the same conditions as those under which the trainee will be tested.

The pilot should ideally have two assessors to monitor the performance of a single individual. This should be a simultaneous but independent assessment. If the scores (satisfactory/unsatisfactory) are different for any of the steps, a reliability problem exists. When conducting the pilot, the assessors should look for problems or deficiencies such as:

— Questions asked by the trainee;
— Equipment requirements;
— The ability of the trainee to perform the task;
— Unclear directions to the trainee;
— Unusual conditions or problems beyond the assessors' control that affect the outcome of the test;
— The effectiveness of the scoring method used;
— Time considerations.

After the performance test has been piloted, reviewed and corrected from feedback, it should be approved. The test package should be signed and dated by both facility and training representatives.

6.2. PERFORMANCE TEST ADMINSTRATION

Test administration has an important effect on the usefulness of test results and requires control. The assessor should ensure that a suitable environment is established, clear test directions given and that proper supervision is provided for the entire test.

6.2.1. Establish environment

Effective testing requires that the physical qualities of the test environment and setting within which the trainee performs are satisfactory. High noise levels, poor lighting, lack of ventilation, excessive heat or cold, adverse safety conditions and frequent interruptions will lower trainee test performance. Prior to the performance test, the instructor should ensure that the conditions of the test location are adequate.

6.2.2. Test directions

Prior to conducting a performance test the assessor should provide the trainee with directions and an overview of the performance testing process. These directions should provide the trainee with clear and complete instructions as to what the trainee will be allowed to do and when the instructor will allow the trainee to do it. The assessor should explain under what circumstances he/she will stop the trainee if concerns such as safety of personnel or equipment arise.

6.2.3. Conducting the performance test

Completion of the task is not the only indicator of the competence level of the trainee. It is important to observe the methodology as well as the outcome of the performance test. Some typical questions that the assessor should consider when observing a performance test include:

(a) Were the tools used correctly and in the proper sequence?
(b) Were the necessary reference materials obtained?
(c) Were non-critical steps performed in the proper order?
(d) Was the trainee confused by any portion of the performance test?
(e) Was the equipment manipulated in a deliberate and timely manner?
(f) Was the trainee aware of equipment status (e.g. did he/she recognize when a pump was running or when a valve was open)?
(g) Were safety rules observed when performing the task?
(h) Was the appropriate personal protective equipment worn (if applicable)?

Complete testing of knowledge and skills requires the assessor to question the trainee during the performance test. However, the assessor should not ask distracting questions; all questions should be related to the task. The assessor may ask the trainee to 'talk through' the task as he/she performs it. This technique reduces the number of questions the assessor needs to ask and allows the instructor to stop the trainee before he/she makes a serious mistake. The questions may be written in the assessment standard (preferred method) or generated by the assessor during the performance test.

During the conduct of a performance test, the assessor must also act as safety monitor in addition to performing his/her role as evaluator. The assessor has responsibility for stopping the performance test whenever personnel injury or equipment damage can occur, public health or safety is affected, or the trainee deviates from an approved procedure.

6.2.4. Assessing the performance test

Scoring methods should be identified and should be closely related to the evaluation standards. Trainees should be evaluated by qualified assessors on how closely their performance meets the standards. Some rating method examples are pass/fail, satisfactory/unsatisfactory, yes/no and 80% correct.

6.2.5. Debriefing the trainee

At the completion of a performance test the assessor and the trainee should conduct a detailed review of the trainee's performance. The assessor should tell the trainee if he/she passed or failed the performance test. The review should be conducted immediately while the events are fresh in the minds of the assessor and the trainee. The instructor is responsible for recording the results accurately before, during and after the performance test. Accurate recording of results allows the testing process to be evaluated, ensures fair grading and allows for monitoring the results to ensure reliability.

6.3. OTHER TYPES OF PERFORMANCE ASSESSMENT

6.3.1. Assessment centres

Assessment centres are sometimes established within a training organization and are designed to assess the supervisory/management potential of individuals. They utilize simulated workplace situations in which the trainee is required to demonstrate his/her competency to perform specified activities which are complex in nature and representative of tasks for the higher level position. These activities are designed to challenge trainee competency and to act as a predictor of success in the higher level position.

Examples of typical assessment centre activities include establishment of a simulated office environment in which the trainee is required to handle multiple tasks under time pressure conditions, or to engage in strategic and financial planning exercises, where information may be ambiguous or incomplete and a high degree of judgement is needed. Senior managers are often used to observe and assess the performance of the trainee, using similar principles to those described earlier in the section, but drawing heavily on their knowledge and experience of the organization. Annex B in Ref. [6] contains a description of assessment centre actvities and the accompanying CD-ROM also has severval examples of assessment methods used to recruit and select personnel.

6.3.2. Performance assessments of professional–technical staff

A signicant portion of the training of professional and technical staff such as engineers and instructors is accomplished through mentoring and coaching. At the completion of mentoring, trainees are evaluated by their mentor to determine if they have mastered the performance objectives. The same principles disccussed in Sections 6.1 and 6.2 for the assessment of performance are used for professional and technical staff. These assessments typically combine oral and performance examinations. However, the assessment forms are not usually as detailed as OJT or JPM assessment instruments. Typically, this instrument consists primarily of a discussion or interview in which the mentor uses professional judgement to determine if the trainee has achieved mastery of the performance objectives. The assessment can also include mentor observatioin of trainee activities. Discussion questions are usually included in the assessment. Sign off is formaly recorded when the trainee has demonstrated competency.

The appendices contain several examples of assessment forms. Appendix XIII is an example of an assessment form for a systems engineer, Appendix XIV is an example of an assessment form for an instructor and Appendix XV is an example of a self-assessment form completed by the instructor.

6.3.3. Simulator assessments

This topic is addressed in the IAEA technical document Use of Control Room Simulators for Training of Nuclear Power Plant Personnel [7].

7. COMPUTER BASED EXAMINATIONS

Computer based testing can be used as an assessment tool as part of a conventional training course, or as an integral part of a distance, web based or e-learning training course. Increasing use is being made of e-learning platforms to implement training programmes. There are many advantages offered by employing the computer environment in the assessment process. Some of the advantages of computer based assessment over traditional paper and pencil testing are:

(a) Savings in money and time (no paper, answer sheets, copying hand grading, etc.);
(b) Capability to generate random tests;
(c) Capability to generate individualized tests;
(d) Capability to provide real time training/testing outside normal working hours;
(e) Capability to provide training/testing from remote locations;
(f) Improved examination security and reduced opportunity for cheating;
(g) Achieves immediate results;
(h) Provides feedback to trainee with links to learning resources;
(i) Allows comparison of test results over time;
(j) Automated test item analysis tools to improve reliability and validity;
(k) Provides for pre-testing and test out;
(l) Provides a convenient means for practice and self-study.

However, there are some limitations and concerns with regard to using computers for assessments. One of the significant disadvantages is the inability of computerized assessments to make judgements on the personality and attitude of the trainee. The capability to evaluate essay type answers is also extremely limited at this time. Also, the absence of a relationship and direct interaction with the instructors does not allow the instructors to follow the evolution of their competencies. Nevertheless, the development of the e-learning applications is addressing these limitations and will improve the use of computer based assessments.

The use of multimedia technology also makes it possible to design more realistic tests by using simulation methods involving full colour, motion and sound. A broader array of test item formats can be used. Many of these formats are variations of paper and pencil questions and include:

— Drag and drop (moving a label to the correct component);
— Hot spot (clicking on correct part of a diagram);
— Matrix (allow selection from a matrix list of choices);
— Pull down lists;
— Ranking;
— 'Select-a-blank' (fill in a blank from a pull down menu);
— Short answer questions;
— Simulation type questions.

Computer based testing applications can be used on local area or web based networks, making it possible to provide training and testing in a number of different environments, including a centralized assessment centre, or at

individual work stations. In addition, training and testing can take place outside normal working hours from remote locations. All of the guidance in Section 4 is applicable to computer based tests.

The accompanying CD-ROM contains several examples of computer based testing assessment procedures and tools.

8. EXAMINATION METHODS TO IMPROVE THE EFFICIENCY AND EFFECTIVENESS OF TRAINING

While the computer based examinations discussed in Section 7 can improve the efficiency and effectiveness of examinations, other methods can also be implemented to improve the overall efficiency and effectiveness of the resources devoted to examinations. This section briefly discusses a few approaches that are being used by some Member States.

8.1. SOME MEMBER STATE APPROACHES

8.1.1. Sharing of training and examination materials

For many job positions, common core of training material, including examination questions, can be used. For example, for job positions such as radiation protection technicians and maintenance crafts, a significant portion of initial training is the same even for different types of nuclear power plant. In some Member States this training material is made available through the internet, either openly or through password controlled sites. The use of this material can significantly reduce the costs of development of new material. One example of such training material for radiation protection technicians is provided by the United States Department of Energy (http://www.eh.doe.gov/radiation/RST/rst.htm). This site also contains training material and examination banks for radiation worker training and for general employee training on radiation safety.

The IAEA has also established a programme to support the development of standardized training packages and distance learning tools in the field of nuclear safety. Information on this programme and the training materials available can be found on its web site (http://www-ns.iaea.org).

8.1.2. Sharing of examination banks

In some Member States, nuclear industry organizations provide examination banks on the internet. An example of such examination banks for pressurized water reactors and boiling water reactors is provided by the Nuclear Regulatory Commission (http://www.nrc.gov/reactors/operator-licensing.html). The accompanying CD-ROM has several examples of examination banks.

8.1.3. Industry-wide examination process

In some Member States, industry groups develop common examination material for selected nuclear industry job positions. One example of such examination material is the task performance evaluation programme developed by the Electric Power Research Institute. This programme is an industry-wide accepted process for examining, registering and tracking maintenance workers' competency on a variety of common maintenance tasks (e.g. valves, pumps, circuit breakers, transmitters, rigging and hoisting). This programme provides the capability to pre-qualify contracted maintenance personnel who are hired during plant outages.

8.1.4. Reciprocity agreements

A reciprocity agreement is a process that allows an organization to accept previous training and qualification of personnel from another, similar organization. It can be used for almost any job position, but it is most widely used for general employee training, including training for outage support contractors, who may go from one plant to another for outage work.

8.1.5. Nuclear training groups

In a number of Member States, nuclear power plant training centres have joined with other nuclear power plant training centres to promote the sharing and exchange of training information and materials as a means of improving the effectiveness and efficiency of training and assessment. Examples include the IAEA sponsored Pan European Training Association and various US nuclear training associations, such as the Middle Atlantic Nuclear Training Group (MANTG). Most of these groups maintain internet sites for the sharing of information. Examples can be found on ENTRAC (http://entrac.iaea.org), Asian Nuclear Safety Network (http://www-ansn.iaea.org), MANTG (http://

www.mantg.com/), Western States Training Group (http://www.westrain.org/) and MidWest Nuclear Training Association (http://www.ncweb.com/org/mnta/).

9. RELIABILITY AND VALIDITY

Because tests are used to qualify personnel to do a job or task, it is important that they are developed properly. If tests are constructed systematically and administered correctly, they will have a high degree of reliability (i.e. consistent results). The quality and effectiveness of tests should be monitored continually and improved where necessary. Analysis of test results provides important input to the quality and effectiveness of tests. Although most instructors and test developers are not required to perform complicated statistical analyses, an understanding of some basic concepts is beneficial in interpreting and refining the testing process.

This section provides a brief discussion of reliability and validity. Readers are encouraged to consult the Bibliography and other available textbooks on these subjects.

9.1. RELIABILITY

Reliability is functionally defined as the consistency between two separate measurements of the same item. If a test gives perfectly consistent results, it would be perfectly reliable. Reliability is generally not a problem with performance tests as long as conditions in the assessment situation remain constant. Reliability can be a problem with written tests because test item construction can be difficult. Reliability can be affected by ambiguous test items, multiple correct answers, typographic errors, adverse testing conditions, interruptions, limited time and complicated answer sheets. Trainee readiness and scoring errors also affect test reliability.

Four methods which can help ensure that an examination is reliable are: (1) test/retest, (2) alternate form, (3) split half and (4) inter-item correlations.

(1) Test/retest. This method involves administering the same test or survey to the same group of employees at two different time periods and calculating the correlation of the scores. If there is a high degree of positive correlation, then the test is reliable.

(2) Alternate form method. This method involves constructing two similar tests and administering them to employees at the same time and analysing the correlation between the two scores. If there is a high positive correlation, then the test is considered to be reliable. Constructing a similar test can be time consuming, which may make this approach difficult.

(3) Split half method. This method involves splitting the test into two equal parts and comparing the results. For example, it might be appropriate to compare the even numbered questions with the odd numbered ones. The scores of the two halves are compared and their correlations checked. A high correlation indicates a reliable test.

(4) Inter-item correlations. A fourth method to measure reliability is to calculate correlations between each of the items in the test. For example, a test with 25 items is divided into 25 parts. A correlation is developed comparing each item with all of the others.

The following examples illustrate how reliability or unreliability may be indicated as tests are analysed.

Example 1. Ten trainees were given test A on Monday and then again on Tuesday. Assuming that nobody forgot anything overnight, the Tuesday test results should be exactly the same as the Monday test results if test A is reliable. Any significant difference would indicate test unreliability since nothing changed from Monday to Tuesday.

This is a form of test/retest reliability. The time period for this type of reliability is variable. Longer time periods generally result in greater differences in test results, but long time periods can determine the long term stability of the test or indicate lack of knowledge retention.

Example 2. Ten trainees took a test and nine of them missed question #5. Question #4 was not missed by anyone but was testing knowledge very similar to that tested by question #5.

Question #5 may be unreliable due to poor wording, unclear answers, a typographic error that makes a wrong answer look correct, etc. This is a form of alternate question reliability.

Example 3. Eight of ten trainees who missed question #7 chose answer (b). Does answer (b) look too similar to the correct answer? Does the lesson plan support the correct answer?

The above example could indicate the use of a method of testing known as key word and tricky phrase testing. This type of testing encourages the trainee to memorize and recall only key words and tricky phrases to pass the test instead of requiring the trainee to learn the material. It is, therefore, a poor method to use.

Test items having poor reliability are easy to recognize. If trainees of equal knowledge or ability have widely varying test scores, the test or test item may be unreliable. Also, if the same trainee is tested twice on the same test or test item within a short period of time and passes once but fails the next time, the test or test item may be unreliable. In both of these cases the reliability should be questioned and the test or test item carefully re-evaluated.

9.2. VALIDITY

A valid test measures exactly what it was intended to measure. A test can be reliable but not valid. A paper and pencil test can be reliable in measuring knowledge of certain welding fundamentals but not valid for measuring welding skill. Establishing the validity of tests can be a complicated and time consuming process. Validity can be improved by:

(a) Ensuring that a good analysis of the job has been conducted;
(b) Ensuring that knowledge, skills and, as applicable, attitudes have been identified;
(c) Ensuring that training objectives for both knowledge and skills are based on task requirements;
(d) Identifying the type of performance dictated by objectives (cognitive, psychomotor, affective);
(e) Ensuring that action verbs used in objectives measure what they are intended to measure;
(f) Designing test specifications to ensure that objectives are covered adequately;
(g) Discussing the test with SMEs, supervisors and training specialists;
(h) Piloting the test or sample test items with SMEs and trainees;
(i) Comparing test results with actual job performance;
(j) Ensuring that the test and test items are changed in order to be consistent with revised job requirements.

Because reliability is a necessary component of validity, it should be established before attempting to assess validity. There are four approaches used to determine if an examination is valid. These approaches, adopted by the American Psychological Association, are: (1) content validity, (2) concurrent validity, (3) construct validity and (4) predictive validity.

9.2.1. Content validity

Content validity is the simplest method to employ in order to assess whether a test is valid. Content validity can be established by comparing the test items with the training objectives. No statistical calculations are needed to establish content validity. If SMEs agree that the test items measure their respective training objectives, the test can be considered valid. The usefulness of content validity is subject to the quality of the analysis and the subsequent training objectives as well as to the thoroughness of the SME review of the test items. A content valid test should sample all important areas of performance and sample these areas in approximate proportion frequency, importance and complexity as they occur on the job. The number of test items in an examination should also correspond with the amount of time, exposure, or importance of the material presented.

9.2.2. Concurrent validity

Concurrent validity refers to the extent to which one test compares favourably with another, already validated test. If there is already a valid measure (i.e. a nationally recognized entrance examination) of what is to be tested, determine the degree of association between the results of the pre-established test and the test to be validated. To the extent that they are related, there is an established level of concurrent validity. A statistical analysis is required to establish a level of concurrent validity. Information on statistical analysis to determine concurrent validity can be found in many available textbooks on statistics.

9.2.3. Construct validity

Construct validity refers to the extent to which a test represents the construct it purports to measure. A construct is the underlying concept or variable such as the skill, attitude, or ability that the test is intended to measure. Examples of constructs are:

— Attitude towards supervisor;
— Ability to read a scale;
— Skill in conducting an effective performance discussion.

As a first step in demonstrating construct validity, define all parts of the construct and then show that the test is an adequate measure of that construct.

The definition of the construct should be as clear as possible in order to make it easy to understand. Construct validity can be demonstrated in several ways:

— Expert opinion;
— Correlations;
— Logical deductions;
— Criterion group studies.

Expert opinion is a relatively easy approach. A group of experts state that the test, in their opinion, is an accurate measurement.

Correlations are more complex. In this case another test is used to measure the same or a similar construct and the results compared with the first test. Positive correlations could indicate validity.

Logical deduction is more subjective. The test must logically conclude, through a series of deductions, that the instrument actually represents a measure of the construct.

The criterion group may be more useful. A group of people possessing a deficiency of the construct in question has the test administered to them. If, indeed, the results agree with the existing knowledge about the group, then it helps make the case for construct validity.

9.2.4. Predictive validity

Predictive validity refers to the extent to which a test can predict future performance after a given time interval has elapsed, e.g. when trainee scores on one test can be used to predict success on a second test administered at a later time. Establishing predictive validity is accomplished in a similar manner to establishing concurrent validity. Statistical analysis is used to determine predictive validity as long as both tests are scored on a pass or fail basis and the tests are separated by a substantial period of time.

9.3. METHODS TO IMPROVE VALIDITY

There are no magic formulas to ensure that a test is valid when it is designed. However, there are a few simple guidelines that may help improve validity:

(a) Include a suitable number of appropriate items. Too few items in an examination can hamper the validity, while too many can be cumbersome and time consuming. Strive for the right balance to improve the validity.

(b) Reduce response bias. Participants responding to questions on an examination may tend to say what they think the instructor wants them to say. This desire to please can make an examination invalid. Participants should be encouraged to give candid responses.
(c) Be objective in administering the examination. In some cases the staff administering the examination may be biased in the outcomes they expect. For instance, if it is revealed that one group is expected to outperform another, it can sometimes influence the results. The participants may give responses that suggest improvement when actually those improvements do not exist.
(d) Recognize the weak link between attitude and behaviour. Attitudes do not always predict behaviour. A person may indicate certain attitudes but behave in a different manner. Even the most carefully designed examination can miss in a prediction of behaviour or performance.

10. EXAMINATION AND TEST ITEM ANALYSIS

This section provides a brief description of several typical measures or statistics that can be used for the interpretation and evaluation of examination results and the analysis of individual test items. Readers are encouraged to consult Refs [12, 13], the Bibliography and other available textbooks for methods used to calculate the measures discussed in the following subsections, as well as examples of their uses for interpreting examination results and individual test items. The References and Bibliography also identify many other measures that can be useful in interpreting examinations and test items. Many testing software applications are also available that provide for automatic generation of examination and test item statistics.

10.1. EXAMINATIONS

There are a number of useful methods that can be used to interpret the results of examinations. These include:

(a) Raw score. The actual score a student receives on a test. Raw scores need to be evaluated in relation to the grading scale used on the test.

(b) Frequency distributions. A display of test scores arranged from highest to lowest. Frequency distributions are useful to assess how well one trainee did in comparison to other trainees and can also provide useful information to management on the examination results of all trainees.
(c) Ranks and percentiles. The rank indicates the position of a score in relation to other scores. Percentile rank can also be calculated to indicate the percentage of trainees in a class who obtained raw scores below a certain value. These measures help clarify the meaning of raw scores.
(d) Measures of central tendency. These measures include the median, the mean and the mode.
(e) Median score. The middle score or 50th percentile score. The median score divides the test results into two equal parts.
(f) Mean score. The average score in a group of scores on a test. The mean is computed by adding all the scores and dividing the sum by the number of scores.
(g) Mode. The most frequent score in a frequency distribution.

The following measures also provide useful information for training management as well as to the trainees:

(a) Measures of dispersion. These measures include the range and the standard deviation.
(b) Range. The difference between the highest and lowest scores on a test. The range is used with measures of central tendency such as the mean to identify the variability of test results between two groups of trainees on the same test.
(c) Standard deviation. A measure of the variability or dispersion within a set of scores. It indicates the extent to which scores vary around the mean or average score.
(d) Z scores. The number of standard deviations a raw score departs from the mean score in a distribution.

Methods for calculating the above measures and examples of their use for interpreting test results can be found in Refs [12, 13] and many of the texts listed in the Bibliography.

10.2. TEST ITEM ANALYSIS

In addition to evaluating and interpreting overall examination results, there are a number of measures that are useful in interpreting and improving individual test items. These include:

(a) Response profile. An index of the frequency of trainee responses to test answer alternatives. It is most useful in the identification of how well distractors work in multiple-choice questions discussed in Section 4.3.
(b) Item difficulty. The proportion of trainees who responded correctly to a test question. It is expressed on a scale from 0.0 to 1.0. The lower values indicate difficult questions; higher values indicate easy questions. The value is usually between 0.3 and 0.7.
(c) Item discrimination. An index of how well a test question discriminates between trainees who scored high and low in an examination. It is expressed on a scale from +1.0 to – 1.0. Positive discrimination means that more high than low scoring trainees responded correctly to a test item. Negative discrimination means that more low than high scoring trainees responded correctly to a test item. No discrimination means that an equal number of high and low scoring trainees responded correctly to the test item. In the latter two cases, the test item is faulty and needs to be revised or discarded.

The above measures are typically referred to as traditional or classical test item analysis. These traditional measures provide an effective method for the review and evaluation of individual test items. The methods for calculating the above measures and examples of their use for interpreting test items can be found in Refs [12, 13] and many of the texts listed in the Bibliography. Readers interested in more detailed measurement theory, such as item response theory or Rasch measurement should consult the Bibliography for texts on these subjects. The accompanying CD-ROM also has several examples of administrative procedures for test item analysis.

Appendix I

DEFINITION OF TERMS*

Ability. The mental or physical power or talent to undertake an activity, either innate or acquired, through learning, practice and undergoing training. Ability encompasses attitudes, knowledge and skills.

Action verb. A word that conveys action or behaviour and reflects the type of performance that occurs (e.g. to hold, open, describe, calculate, justify). Action verbs reflect behaviours that are both observable and measurable.

Affective. Relating to or resulting from attitudes, emotions, values or feelings (see affective domain and attitudes).

Affective domain. One of three areas used to classify learning objectives/training objectives containing those relating to attitudes (feelings, perceptions and values). Also known as the attitudes area. The accepted taxonomy (ascending order or level of complexity) within the affective domain is:

— Attending: Pays attention to received stimuli or events.
— Responding: Reacts positively to stimuli or events by participation.
— Valuing: Demonstrates belief in the worth or value of an event or activity.
— Organization of values: Compares various values and prioritizes them.
— Characterization by values: Displays an attitude characteristic of a pervasive, consistent and predictable set of values.

Analysis. (1) The fourth level in the cognitive domain or knowledge area that involves breaking down an idea into its constituent parts and examining their interrelationships. (2) A method of subdividing a problem to be able to make decisions; examples are algorithms, network analysis, critical path analysis.

Analysis phase. The initial phase in the systematic approach to training that serves as the foundation for training programme design, development, implementation and evaluation. The analysis phase assesses performance requirements or deficiencies in order to be able to identify the competences, in terms of knowledge, skills and attitudes, needed.

Aptitude. The ability to learn when given the opportunity and suitable training, also the inherent (or learned) ability to do something.

Assessment. A structured activity by which the knowledge and/or skills and/or attitudes of an individual are measured using one or more methods. The

* Modified from IAEA-TECDOC-1358 [14].

exact purpose of assessment (confirming competence, predicting future performance, etc.) determines which assessment method is used. Assessment is often conducted at the end of a training session or course to determine the extent to which trainees have met the training objectives.

Assessment fidelity. The extent to which an assessment reflects the achievement of associated training objective(s). The closer the relationship the higher the fidelity of the assessment. Also termed test fidelity.

Assessment method. A method of assessing an individual or group. A multiple-choice question, essay-type question, oral question, assignment, project, quiz, walkthrough and observation are some typical assessment methods. See assessment.

Assessment reliability. The extent of the consistency with which an assessment produces the same results under different but comparable conditions, e.g. each time it is used. Assessment reliability is also termed test reliability. See assessment.

Assessment type. The nature or type of assessment, determined largely by the purpose of the assessment. See assessment.

Assessment validity. The validity of an assessment, but the validation criteria depend on the purpose of the assessment. Assessment validity is also termed test validity. See assessment and validity.

Assessor. An individual assigned to conduct an assessment. See assessment.

Attitudes. The observable characteristics of individuals resulting from their personal emotions, values and feelings that determine ways in which they interact with others and their work and so affect their interpersonal relationships and approach to their job and safety issues. Together with knowledge and skills, attitudes provide the full requirements to perform a given job or task competently. Attitudes are sometimes termed affective abilities.

Bloom's hierarchy *or* taxonomy. The classic example of a learning taxonomy.

Checklist. In training, a listing of tasks or task elements that is used to confirm that these activities have been performed. If the tasks have to be performed in a particular order, this appears on the checklist. See checkout.

Checkout. A type of examination that may be an oral or performance assessment or a combined oral and performance assessment. For example, completion of a checklist is commonly referred to as a checkout. See checklist.

Cognitive domain. One of three areas used to classify learning objectives, containing those relating to knowledge based mental processes. Also known as the knowledge area. The accepted taxonomy (ascending order or level of complexity) within the cognitive domain is:

— Knowledge: Recognizes and recalls information.
— Comprehension: Interprets, translates or summarizes given information.
— Application: Uses information in a situation different from the original learning context.
— Analysis: Separates wholes into parts until relationships are clear.
— Synthesis: Combines elements to form a new entity from the original one.
— Evaluation: Involves acts of decision making based on criteria or rationale.

Competence (competency). (1) The ability to put skills, knowledge and attitudes into practice in order to perform activities or a job in an effective and efficient manner within an occupation or job position to identified standards. (2) A combination of knowledge, skills and attitudes in a particular field, which, when acquired, allows a person to perform a job or task to identified standards. Competence (competency) may be developed through a combination of education, experience and training.

Comprehension. The second level in the cognitive domain or knowledge area.

Criterion. A characteristic or measurement with which other characteristics or measurements are compared, usually being a standard against which something is measured. In training, the task or training objective standard is a measure of trainee performance. In test validation, it is the standard against which test instruments are correlated to indicate the accuracy with which they predict individual performance in a specific area. In assessment it is a measure used to determine the adequacy of a performance or behaviour. In evaluation it is the measure used to determine the adequacy of a product or process. A criterion in the training context is sometimes termed a standard.

Difficulty. A dimension used in rating tasks or prioritizing training objectives that reflects how difficult it is to learn or perform a task.

Enabling (training) objective. A statement of intent, especially the expected outcome of a segment of training. An enabling objective must include the expected performance and state or imply the associated conditions and standards. Meeting an enabling objective helps a trainee attain one or more terminal objectives or training aims.

Entry level requirements. The identified levels or standards of a combination of education, training and experience required to enter a training course, module or programme.

Entry level test. An assessment containing items based on the prerequisites that the intended trainees must have mastered in order to begin a training course, module or programme.

Examination. An assessment in the form of a formal series of questions or tests which trainees must complete, usually in a fixed time and normally under controlled conditions, to ensure there is no unauthorized collaboration. Examinations are often administered at the conclusion of a training course or programme. Less formal tests take place during or after training sessions and lessons.

Job analysis. A method used to obtain a detailed listing of the duties and tasks of a specific job. See task analysis.

Job and task analysis. A combination of job analysis and task analysis. See job analysis and task analysis.

Job competency analysis. A type of analysis that concentrates on the education, experience and competence required to perform a job to required standards. See competence and job analysis.

Job performance measure. A test used to assess the level of performance of a job incumbent or trainee on a specific task, or set of tasks, against predetermined standards.

Knowledge. (1) The mental constructs used in acquiring and understanding facts and the application and reassembly of facts to think creatively, solve problems and make judgements. Together with attitudes and skills, knowledge provides the full requirements to undertake a given job or task. Knowledge is sometimes termed cognitive ability. (2) The lowest level in the cognitive domain. See cognitive domain, attitudes and skills.

Knowledge area. An alternative name for cognitive domain.

Learning objective. A precise specification of what behaviour is to be learned in terms of the expected performance, the conditions under which the performance is demonstrated and the standards or level of performance. The two types of learning objective are terminal objectives and enabling objectives. Learning objectives are the same as training objectives except that training objectives focus on the training that is to be provided, while learning objectives focus on what an individual is to learn.

Learning taxonomy. A classification of cognitive, affective and psychomotor behaviours in three taxonomies (hierarchical orders or levels of ascending complexity). Devised by B.S. Bloom and his colleagues to interpret teaching, learning and assessment and applied in many training environments. Three domains or areas are identified: affective domain (attitudes area), cognitive domain (knowledge area) and psychomotor domain (skills area). Also termed Bloom's hierarchy or taxonomy.

Objective test. A test or assessment in which subjective bias is eliminated by providing the answers to questions as fixed choices. The answers, therefore, require no qualitative interpretation and can be marked with reliability by non-subject specialists or electronically, e.g. by a computer.

Objectives. In training, a term that is applicable to, and incorporates, enabling objectives and terminal objectives.

Open-ended. A term applied to a question, test, exercise or project that has more than one acceptable outcome or response, rather than one correct solution or result. An essay-type question, for example, is open-ended.

Performance. The display or achievement of ability in undertaking a specific activity. In training, the conditions of performance and the standards required are normally specified. Performance is the main output used in the behavioural approach to learning or training, where the actual performance is compared with the expected, pre-specified performance under stated conditions and standards. See ability.

Performance based training. Training based on mastery criteria in which the relevant knowledge, attitudes and skills required for competent job performance have to be demonstrated by the trainee. The systematic approach to training is one example of performance based training.

Performance objective. In training, an alternative term for training objective.

Performance test. An assessment of performance. This may involve a practical demonstration, by a trainee, of the ability required to perform a task that is assessed by a qualified instructor.

Pre-test. A test carried out before an activity (also known as pre-assessment).

Proficiency. The ability to perform a specific activity (e.g. a task) to demonstrate mastery of that activity.

Psychomotor domain. One of three areas used to classify learning objectives, containing those relating to physical skills (movement and coordination). Also known as the skills area. The accepted taxonomy (ascending order or level of complexity) within the psychomotor domain is imitation, manipulation, precision, articulation and naturalization (see skills, affective domain and cognitive domain):

— Imitation: Observes a skill and tries to repeat it.
— Manipulation: Performs a skill according to instruction rather than observation.
— Precision: Reproduces a skill with accuracy, proportion and exactness. Usually performed independent of original source.
— Articulation: Combines one or more skills in sequence with harmony and consistency.
— Naturalization: Completes one or more skills with ease and becomes automatic.

Quiz. An informal, often oral, assessment in which the correct answer is provided immediately after the learner has offered a response. A quiz is conducted usually at the conclusion of a training session.

Skills. The physical and manipulative actions following the mental signal needed to perform an activity or task. A term often incorrectly applied to abilities. Together with attitudes and knowledge, skills provide the full requirements to undertake a given job or task to specified standards. Skills are sometimes termed psychomotor abilities. See psychomotor domain, attitudes and knowledge.

Skills area. An alternative name for psychomotor domain. See psychomotor domain.

Soft skills. The practical application of attitudes in performing a task or undertaking a responsibility in a job position. Examples of soft skills are human interactions such as leadership, teamwork, communication, reinforcement, critiquing, assessing, coaching, observing, counselling, supervising and managing.

Subject matter expert. An individual who, by virtue of education, training and/or experience, is a recognized expert on a particular subject, topic, or system, or who is acknowledged as being highly competent in performing a particular task. A subject matter expert may be one of a team of experts.

Task. A measurable, well-defined unit of work, with an identifiable beginning and end. Several tasks, which may be arranged within a duty area, are components of a job.

Task analysis. The formal identification of the knowledge, skills and attitudes that are required to perform tasks associated with a particular job competently.

Terminal objective. In the training context, a statement on the purpose or goal of a particular training session, course or programme. Also termed a training aim or instructional aim. A terminal objective is usually written in behavioural terms, stating the expected outcome in terms of performance, conditions and standards, but it may also be written in general terms, supported by enabling objectives which are always written in behavioural terms. Terminal objectives are intended for long term retention and are reinforced through continuing training as needed.

Test. A method of assessment. See assessment.

Test reliability. The extent of the consistency with which a test produces the same results under different but comparable conditions, e.g. each time it is used. See test.

Test validity. The validity of a test. The validation criteria are determined largely by the purpose of the test. See test and validity.

Validity. In training, the extent to which an item, such as a task statement or qualification, fulfils or represents the purpose for which it was intended. Validity is commonly used in evaluation.

Walkthrough. A method of oral assessment in the trainee's work area where the assessor and trainee 'walk through' or alongside the plant and the assessor asks the trainee questions relating to items of equipment or plant relevant to the trainee's training objectives.

Appendix II

ACTION VERBS

TABLE 8. ACTION VERBS VERSUS LEVEL FOR THE COGNITIVE DOMAIN

Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
define	convert	apply	analyse	arrange	appraise
describe	defend	associate	appraise	assemble	assess
identify	describe	build changes	break down	categorize	choose
label	diagram	calculate	classify	collect	compare
list	discuss	choose appropriate procedures	compute	combine	conclude
locate	distinguish	collect information	conclude	compile	contrast
match	estimate	compute	contrast	compose	criticize
memorize	expand	construct	criticize	construct	critique
name	explain	demonstrate	debate	create	decide
outline	express	discover	determine	design	describe
recall	extend	dramatize	diagram	develop	develop criteria
recite	generalize	employ	differentiate	devise	discriminate
recognize	give examples	find solutions	discern	explain	estimate
record	identify	illustrate	discover	forecast	evaluate
relate	illustrate	interpret	discriminate	form	explain
repeat	infer	locate	distinguish	formulate	interpret
reproduce	interpret	manipulate	examine	generalize	judge
state	locate	modify	experiment	generate	justify
select	measure	operate	generalize	hypothesize	make judgement
tell	outline	perform	identify	manage	measure
	paraphrase	practice	illustrate	modify	rate
	predict	put in operation	infer	organize	revise
	recognize	relate	inspect	plan	score
	reconstruct	repair	inventory	predict	select
	report	schedule	locate	prepare	summarize
	restate	shop	note	produce	support
	review	show	organize	propose	
	rewrite	sketch	outline	rearrange	
	summarize	solve	point out	reconstruct	
	translate	translate	question	relate	
		write	relate	reorganize	
			select	revise	
			separate	rewrite	
			solve	set up	
			subdivide	structure	
			test	summarize	
				tell	
				theorize	
				write	

TABLE 9. ACTION VERBS VERSUS LEVEL FOR THE AFFECTIVE DOMAIN

Awareness	Reinforcement	Promotion	Defence
ask	acclaim	accept	abstract
accept	adhere	adhere	act
accumulate	applaud	advocate	argue
attend	approve	alter	arrest
be aware of	augment	argue	avoid
choose	commend	assist	balance
combine	comply	encourage	debate
control	conform	follow	defend
describe	discuss	help	define
differentiate	follow	initiate	discriminate
follow	obey	justify	display
read	play	model	formulate
receive	practice	prefer	influence
recognize	praise	propose	intervene
reply	volunteer	subsidize	manage
respond		select	organize
select		support	prevent
separate			resist
set apart			resolve
share			

TABLE 10. ACTION VERBS FOR THE PSYCHOMOTOR DOMAIN

acknowledge	compare	heat	override	shut down
activate	complete	heat up	perform	sketch
actuate	compute	hoist	plot	splice
add	connect	hold	position	spray
adjust	control	identify	prepare	start
align	cool	immerse	pressurize	start up
alternate	correct	increase	prime	steer
analyse	construct	inform	print	stop
apply	cut	inspect	pull	store
assemble	couple	install	pump	switch
assess	decrease	insert	push	supply
assist	de-energize	isolate	purge	synchronize
backwash	depress	investigate	rack in/out	tagout
balance	deselect	jog	raise	test
begin	detect	let down	reactivate	throttle
bleed	dilute	line up	read	titrate
block	direct	load	rebuild	trace
boil	disassemble	locate	recirculate	track
borate	disconnect	lock	record	transfer
build	display	lower	regulate	transmit
bypass	dispose	lubricate	release	transport
calculate	dissolve	maintain	remove	trip
calibrate	don	manoeuver	repair	tune
call	draft	manipulate	replace	turn
centre	draw	measure	return	type
change	energize	mix	rinse	unlatch
charge	enter	monitor	run	unload
check	establish	move	sample	unlock
choose	estimate	neutralize	scan	uncouple
circulate	exit	observe	secure	vent
clean	feed	obtain	select	warm
clear	fix	open	sequence	warm up
close	flush	operate	service	weigh
code	formulate	organize	sharpen	weld
collect	guide	overhaul	shut	withdraw
				zero

Appendix III

TEST ITEM DEVELOPMENT FORM

Test item reference number:	
Test item format: • Short answer _____ • Matching _____ • Multiple-choice _____ • Essay _____	
Purpose of test item (i.e. objectives tested, areas covered, content tested): Level of difficulty: Reference to course material (training objective, lecture/lesson plan, references, page numbers): Test item statement (provide complete test item with point values): Test item response (model answer, point value, grading criteria, including partial credit): Alternative acceptable answers:	
Review and approval	Date
Developed by:	
Reviewed by (SME):	
Reviewed by (instructional technologist):	
Approved by:	

Appendix IV*

WRITTEN TEST ITEM REVIEW CHECKLIST

The following checklist presents points that should be incorporated when constructing test items from the learning objectives for the written examination.

General guidance

- Does the test item match the learning objective? Does each test item measure one specific concept and reflect the actions, conditions and standards of the objective being tested?
- Is the test item clear, concise and easy to read? Does the test item contain only information relevant to the problem posed? Could it be stated more simply and still provide the necessary information? Can it be reworded or split up into more than one test item?
- Does the test item provide all necessary information, conditions and assumptions needed for a fully correct response?
- Is the test item written at the highest appropriate level of knowledge or ability for the job position of the trainee being tested?
- Is the test item grammatically correct?
- Is the test item free of tricky (deceptive, misleading) wording and clues to the correct answer?
- Is the test item free of unnecessary difficulty or irrelevancy?
- Is the reading level appropriate for the target audience?
- Is the test item limited to one concept or topic?
- Are directions for completing individual test items necessary for clarity? Do they repeat standard instructions provided to trainees at the outset of testing?
- Does the test item have face validity?
- Are key points underlined?
- Is each test item separate and independent of all other test items?
- Are all parts of the test item on the same page?
- Are all required materials, drawings and accompanying test items clearly identified? Can the trainee easily locate them?
- Is there sufficient space provided for answers?
- Are the less difficult test items at the beginning of each section?
- Have the test items been reviewed by others?

* Source: USDOE.

Point values

— Point values are specified for whole test items and all subordinate parts (if more than one response is required).
— Values are assigned relative to other test items in terms of:

  - Significance of successful performance of associated objective to task performance;
  - Learning difficulty and cognitive level;
  - Number of responses required;
  - Difficulty of problem.

— Specific values are assigned for parts of short answer and essay test items.
— Test items which test the same objective have comparable point values.

Short answer test items

— Is there one, short, definitely correct answer for each test item?
— Does the scoring key follow directly from the test item?
— Are clues to the answer avoided?
— Is the required degree of precision specified? For test items requiring computation, specify the degree of precision expected. Try to make the answer turn out to be whole numbers.
— Are the test item statements simple and direct, without extensive qualification?
— Are the test item blanks the same length regardless of the number of words to be entered?
— Does the test item wording avoid grammatical clues to the correct response such as 'a' or 'an' before the blank?
— Limit the space allocated for each answer to encourage a single word or short phrase.
— For fill-in-the-blank test items, arrange the blanks to be of equal but adequate length.
— For fill-in-the-blank test items, do not omit words that are interdependent.
— For fill-in-the-blank test items, do not make sentences unrecognizable by leaving too many blanks.
— For a completion table, do not have more than six items in one column.
— For a completion diagram, include only necessary features and ensure that components are clearly referenced.

— Compose a detailed model answer, identifying important statements, steps, or parts, and allocate points for each test item subpart.

Multiple-choice test items

— Does the test item have one focused topic, making it something other than a collection of true/false test items? Does the stem clearly express a single problem in a direct manner followed by response?
— Does the test item provide sufficient information to provide a basis for formulating the response?
— Is the test item or problem defined in the stem?
— Are tricky (deceptive, misleading) or irrelevant test items avoided?
— Are the answer options homogeneous, highly plausible and a comparable length?
— Are 'none of the above' and 'all of the above' phrases avoided?
— Is one response clearly the correct or best answer accompanied by three or four distractors?
— Is each test item stated positively, unless the intent is to test knowledge of what not to do?
— Is the test item free of 'specific determiners' (e.g. logical or grammatical inconsistencies, incorrect answers which are consistently different, verbal associations between the stem and the answer options)?
— Are common misconceptions used as distractors?
— Are the answer options of the test items ordered sequentially?
— Is the test item free of trivial distractors? Does the stem include irrelevant, trivial detail or instructive information?
— Are the references, attachments and data included in the stem or identified following the stem? The drawings accompanying the stem may precede, be to the right of, or below responses or distractors as space permits.
— Are the correct responses varied from test item to test item in a random fashion?
— Are there equivalent and/or synonymous options that rule out both options for a trainee who recognizes the equivalence.

True or false (multiple-choice) test items

— Does the test item address only one idea or concept in a declarative statement?
— Is the test item either completely true or completely false?
— Is the test item too long or overly complex with qualifying statements?

— Is the test item free of absolute and indefinite terms such as 'always' and 'sometimes' and free of double negatives?
— Are false statements plausible and not false because of trivial detail?

Matching test items

— Are tricky (deceptive, misleading) or irrelevant test items avoided?
— Is there a clearly correct answer or answers to the test item?
— Are clues to the answer avoided (e.g. grammatical clues, response patterns)?
— Do the directions clearly explain to the trainees the basis on which to make the match and how to indicate their answers?
— Do the directions state whether responses can be used more than once?
— Is each response a plausible answer for each premise?
— Are there more responses than premises if each response can only be used once?
— Are the responses arranged on one page in a logical order?
— Is the test item arranged so that the trainees can mark their answers easily?

Essay format test items

— Is the test item clearly and concisely worded *without ambiguity*?
— Does the test item illicit the correct response and no other response? Does it clearly place bounds on the required response?
— Are point values, expected time to respond, expectations for exact answers or estimates clearly identified?
— Is enough information supplied to allow the knowledgeable trainee to respond correctly?
— Is the test item free of negative statements?
— Is the sample answer constructed to minimize subjectivity in grading? Are all significant requirements clearly identified, each with specific point values? Are alternative acceptable responses also included with the sample answer, where appropriate?

Drawing or labelling test items

— Are clear instructions provided concerning what is to be drawn, labelled, or sketched?
— Are clean, readable illustrations provided with parts to be labelled specifically identified?

— Does the test item require only one type of response per diagram?
— Are spaces provided for answers that require labelling a given drawing?
— Does the test item provide sufficient space for the required drawing/sketch?
— Are the points allocated for all parts of the drawing/sketch?

Appendix V*

EXAMPLES OF MULTIPLE-CHOICE TEST ITEMS

A four option multiple-choice test item (a multiple-choice test item with the same correct answer (key)) can be written in several ways depending on the extent of knowledge being assessed. The following example shows three variations of a multiple-choice test item. The correct answer in each case is rad, but the distractors have been chosen with different objectives in mind:

Variation 1:

In health physics, the unit of absorbed dose is:

(a) volt (b) rad (c) calorie (d) newton

The above set of choices covers such a wide range that discrimination would be extremely easy for those with only a vague understanding of health physics.

Variation 2:

In health physics, the unit of absorbed dose is:

(a) rad (b) rem (c) kerma (d) disintegrations/second

The above set of choices covers such a narrow band of knowledge that only those who have a good understanding of health physics would choose the correct answer.

Variation 3:

In health physics, the unit of absorbed dose is:

(a) counts/second (b) rad (c) half-life (d) rem

The above set of choices covers two bands of knowledge. Counts/second and half-life would be immediately dismissed by those with some understanding of health physics, leaving the choice between (b) and (d) (i.e. one choice from two, not four).

* Source: USDOE.

Development of multiple-choice test items

'All of the above' test items provide inadvertent clues to the trainee. When the all of the above option is the correct response, the trainee need only recognize that two of the options are correct to answer the test item correctly. When the all of the above option is used as a distractor, the trainee only needs to be able to determine that one option is incorrect in order to eliminate this option. 'None of the above' responses should not be used with 'best answer' multiple-choice test items, since it may always be defensible as a response.

Each test item should be focused on one topic. A test item containing answer options related to many separate issues does not increase the efficiency of the test item. On the contrary, test items with multiple topics only confuse the trainee about the meaning and purpose of the test item.

The task or problem should be defined in the stem of the test item. Include as much information about the problem or situation in the stem as necessary, leaving only the solution, action, or effect for the answer options. For example:

Poor:

At 50% power:

(a) The equilibrium xenon reactivity worth is approximately equal to the equilibrium xenon worth at 100% power.
(b) The equilibrium xenon reactivity worth is approximately one-half the equilibrium xenon worth at 100% power.
(c) The equilibrium xenon reactivity worth is approximately two-thirds the equilibrium xenon worth at 100% power.
(d) The equilibrium xenon reactivity worth is approximately three-fourths the equilibrium xenon worth at 100% power.

Better:

Which of the following approximates the 50% power equilibrium xenon reactivity worth to the 100% power equilibrium xenon reactivity worth?

(a) Equal to.
(b) One-half.
(c) Two-thirds.
(d) Three-fourths.

Negative stems should be avoided because they are unnecessarily difficult to interpret. Negatively stated stems can also lead to double negatives. It is more important to test what something is rather than what it is not. If a negative stem is necessary, highlight the negative word (e.g. not, never, least).

It is very tempting to write negatively stated test items, since they can be constructed by picking three true statements out of the reference material and changing a fourth statement to make it false. However, trainees do not do as well on negatively stated test items, either because they overlook the negative word or because negatively stated test items require trainees to pick an answer or characteristic that is not true, which can be somewhat confusing. In addition, these test items tend to emphasize negative learning. For example, consider the following stem of a multiple-choice test item:

Test item:

During 100% power operation, the feedwater dump valve opens inadvertently. The condensate pumps will not do which of the following?

This stem can be made to read positively as follows:

Test item:

During 100% power operation, the feedwater dump valve opens inadvertently. The condensate pumps will:

(a) Increase flow to maintain feedwater flow rate.
(b) Trip due to a runout condition.
(c) Have no response.
(d) Trip due to low suction pressure.

There are times when a negatively stated test item is unavoidable. However, *never* use a negatively stated stem with a negatively stated answer option:

Test item:

Which of the following indications would <u>not</u> be expected and might indicate an instrument failure?

(a) The CRD 'travel' lamp does not indicate when Group 8 rods are in motion.
(b) Group 7 out motion is prevented past 91.4%.

(c) When you depress the 'CRD travel in' lamp test push button, the 'CRD travel out' lamp comes on.
(d) During a transfer of a group from direct current, hold to auxiliary, when you select 'SEQ-OR' the SEQ-OR lamp is on and the 'SEQ' lamp goes off.

Notice how confusing the option (a) answer is in combination with the test item stem.

Provide sufficient counterbalance in test items with multipart answers. Multiple-choice test items can legitimately contain multipart answer options. However, if the answers contain too many parts and/or too many options for each part, cues indicating the correct answer may be unavoidable. Consider the following example:

Test item:

The reactor coolant system (RCS) is on hot standby with no reactor coolant pumps running. If OTSG pressure is decreased, according to the plant verification procedure, which of the following temperature responses indicate the presence of natural circulation?

(a) T_h increases, T_c remains the same.
(b) T_h increases, T_c decreases.
(c) T_h decreases, T_c decreases.
(d) T_h remains the same, T_c decreases.

The trainee could choose the correct answer (c) without knowing about the T_c temperature response in this situation, since 'T_h decreases' only occurs in option (c).

Notice that two part answers, with each part containing a two option response, provides complete counterbalance, since all contingencies can be covered in four responses. For example:

Test item:

Which of the following is a definition of quadrant power tilt ratio?

(a) Minimum upper detector output divided by average upper detector output.
(b) Maximum upper detector output divided by average upper detector output.

(c) Minimum upper detector output divided by average lower detector output.
(d) Maximum upper detector output divided by average lower detector output.

A multipart test item which is highly recommended is one in which the two part answer options consist of a two level response (e.g. yes/no, off/on) and a reason. For example:

Test item:

Which of the following best describes the behaviour of equilibrium xenon reactivity over core life?

(a) It decreases, because of the increased fuel burnup.
(b) It decreases, because of the decrease in plutonium–xenon yield.
(c) It increases, because of the increase in thermal flux.
(d) It increases, because of the decrease in boron concentration.

When possible, include common misconceptions as distractors. Since the purpose of the examination is to differentiate between competent and less than competent trainees, a good source of test items involves topics in which there are common misconceptions about important knowledge topics. For example, the following test item was based upon a common misconception about loss of subcooling margin:

Test item:

During a LOCA with a resultant loss of subcooling margin, why are the reactor coolant pumps (RCPs) secured?

(a) To prevent pump damage resulting from operation under two phase conditions.
(b) To prevent core damage resulting from separation upon subsequent loss of RCS flow.
(c) To reduce RCS pressure by removing the pressure heat developed by the RCPs.
(d) To remove the heat being added to the RCS by the operating RCPs.

Make all answer options homogeneous and plausible. For example, the choice of answers to the following test item illustrates these characteristics:

Test item:

On a loss of condenser circulation water intake, the upper surge tank, hot well and condensate storage tank will supply sufficient feedwater to allow decay heat removal for approximately:

Poor	**Better**
15 min	8 h
8 h	24 h
48 h	48 h
3 months	72 h

One method of changing the difficulty level of a test item is to vary the similarity among answer options. Develop distractors that are similar enough to be chosen by those who do not meet the testing objective, yet different enough so that they do not test trivial issues or distinctions. If the answer options have a logical sequence, put them in order.

Avoid overlapping answer options:

Test item:

The SPND uses rhodium which decays with a half-life of 42 s. How long will it take for a detector to indicate approximately 95% of an instantaneous power level change?

Poor	**Better**
2–4 min	1–2 min
4–6 min	3–4 min
6–8 min	5–6 min
8–10 min	7–8 min

To prevent giving the correct answer, ensure that the distractors follow grammatically from the stem. For example:

Test item:

During 100% normal power operation a single steam flow element to the steam generator feedwater fails high. The steam generator feedwater control system will cause:

(a) The feedwater valves to increase the steam generator level slightly before returning the level to normal.
(b) Before returning the level to slightly above normal, the feedwater valves to increase the steam generator significantly.
(c) The feedwater valves to increase the steam generator level to the level of a reactor trip.
(d) The feedwater valves to increase the steam generator level slightly and maintain the increased level.

Test item with improved distractors:

During 100% normal power operation a single steam flow element to the steam generator feedwater fails high. The steam generator feedwater control system will cause:

(a) The feedwater valves to increase the steam generator level slightly before returning the level to normal.
(b) The feedwater valves to increase the steam generator level significantly before returning the level to slightly above normal.
(c) The feedwater valves to increase the steam generator level to the level of a reactor trip.
(d) The feedwater valves to increase the steam generator level slightly and maintain the increased level.

Appendix VI

EXAMPLE DIRECTIONS TO TRAINEES FOR A WRITTEN EXAMINATION

(1) If you have any questions concerning the administration or any part of the examination, do not hesitate asking them before starting that part of the test.

(2) After you complete the examination, you must sign the statement on the cover sheet indicating that the work is your own and that you have not received or given assistance in completing the examination.

(3) To pass the examination, you must achieve a grade of 80% or greater; grades will not be rounded up to achieve a passing score. Every question is worth one point.

(4) You may bring pens, pencils and calculators into the examination room. Use black ink to ensure legible copies; dark pencil should be used only if necessary to facilitate machine grading.

(5) Print your name in the blank provided on the examination cover sheet and the answer sheet. You may be asked to provide the assessor with some form of positive identification.

(6) Mark your answers on the answer sheet provided and do not leave any question blank. Use only the paper provided and do not write on the back of the pages. If you are using ink and decide to change your original answer, draw a single line through the error, enter the desired answer and initial the change.

(7) If you have any questions concerning the intent or the initial conditions of a question, do not hesitate to ask them before answering the question. Ask questions of the assessor or of the designated facility proctor only. When answering a question, do not make assumptions regarding conditions that are not specified in the question unless they occur as a consequence of other conditions that are stated in the question. For example, you should not assume that any alarm has activated unless the question so states or the alarm is expected to activate as a result of the conditions that are stated in the question.

(8) Cheating in the examination will result in failure of the examination and could result in more severe penalties.

(9) Restroom trips are permitted, but only one applicant at a time will be allowed to leave. Avoid all contact with anyone outside the examination room to eliminate even the appearance or possibility of cheating.

(10) When you complete the examination, assemble a package that includes the examination questions, examination aids, answer sheets and scrap

paper and give it to the assessor or proctor. Remember to sign the statement on the examination cover sheet indicating that the work is your own and that you have neither given nor received assistance in completing the examination. The scrap paper will be disposed of immediately after the examination.

(11) After you have turned in your examination paper, leave the examination area as defined by the proctor or assessor.

(12) There is a time limit of three (3) hours for completion of the examination. Do you have any questions?

Appendix VII

EXAMPLE FIELD CHECKLIST FOR A MAINTENANCE TASK

Field checklist for maintenance qualification

Name	
Emp. no.	
Designation	
Level	II and III

TYPE	A	GROUP	M-2
USI NO.	33000	CHECKLIST NO.	M-2-1 A
DESCRIPTION	PHT system		

MISSION	Qualify to perform critical mechanical maintenance tasks on critical equipment of PHT system towards achieving its availability as per standards						
CRITICAL EQUIPMENT	PCP and its motor						
CRITICAL TASKS	See Part-I (enclosed)						
EVALUATION METHOD (Tick the applicable box)	Perform SL Simulator C Class Room	P		Simulate	S		
	Plant	PL		Simulator	SL		
	Classroom	C		Others (specify)	O		
REFERENCE DOCUMENT	01	Procedure no.		33112-1, 2 and 3			
	02	Manufacturer's manual					
	03	Commissioning report					
	04	FS/BM/ED/GA/DD/					
	05	Maintenance manual					
PREPARED BY:		ISSUED BY:		ISSUE DATE			
CHECKED BY:		APPROVED BY:		APPROVAL DATE			
IN ASSOCIATION WITH CORPORATE MPD GROUP							

PART I: PERFORMANCE CHECK

No.	DESCRIPTION	CHECKED BY: (signature and date)
1	Perform the preventive maintenance as per schedule and check for pump and motor alignment.	
2	Prepare a plan for replacement of given PCP motor and execute.	
3	Prepare a plan for overhaul of PCP pump and execute.	
4	Prepare a plan for replacement of PCP mechanical seal and execute.	
5	Prepare a plan for carrying out centring of upper journal bearing.	
6	Perform replacement of the SW gasket on process water side of HP cooler.	
7	Prepare a plan to perform PHT pump motor load balancing and execute.	
8	Prepare a plan for the inspection of motor upper thrust bearing and anti-reverse rotation device when motor is isolated from electrical and oil supplies and execute.	
9	Prepare a plan for the removal of motor coupling half for defective motor and install replacement motor and execute.	
10	Prepare a quality plan for overhauling of PHT motor top bearing replacement and execute.	

PART II: KNOWLEDGE CHECK

II.1. Declared knowledge

No.	DESCRIPTION	CHECKED BY: (signature and date)
1	List complete specification of the given (PCP) pump.	
2	List the protection on this pump and motor.	
3	List all parts of pump and indicate their purpose.	
4	Explain the schematic of mechanical seals. Explain use of backup seal. State material used for seal faces and elastomers.	
5	Describe locking of impeller and wear rings.	
6	Explain the purpose and function of cyclone separator and HP cooler.	
7	State the requirement of the minimum lift of the pump for coupling.	
8	Explain the type of gasket used on pump body and seal housing body.	
9	State the purpose of flywheel mounted on the motor.	
10	Show your understanding on internal construction of thrust bearing of pump and thrust bearing of motor.	
11	State the purpose of oil supply unit to pump.	
12	State the permissible bearing temperature.	
13	Explain gland sealing circuits during: — Normal operation; — Loss of gland injection.	
14	List parameters that have been provided in control room to monitor PCP performance.	
15	State the material used for the PCP: — Impeller; — Wear rings; — Auxiliary impeller; — Mechanical seal faces and thrust bearing; — Heat barrier element.	
16	Explain the consequences of one PCP trip during normal operation.	
17	State the run down time for the PCP.	
18	Describe suspension of pump and tell load distribution.	
19	Explain alignment activities of the PCP. State prerequisite for it.	
20	Explain features and workings of the pressure reducing device on the mechanical seals of the PCP.	
21	Discuss the emergency shutdown and alarm points of the PCP.	
22	State the safety precautions to be taken during overhaul of the PCP motor.	
23	Explain the workings of the oil seal unit circuits.	

II.2. Procedural knowledge

No.	DESCRIPTION	CHECKED BY: (signature and date)
1	Locate and identify the 10 pipelines connected to the seal housing of this pump.	
2	Explain plan for replacement of the mechanical seal of a PCP.	
3	List the criteria for replacement of the seal. State the acceptance standards.	
4	Explain the plan to clean the PCP cyclone separator.	
5	Explain the procedure for replacement of the PCP motor.	
6	Explain the procedure for load balancing of the PCP and its motor.	
7	State the precaution to be taken during load balancing.	
8	State the procedure for overhaul of the PCP pump.	
9	State the procedure for lube oil flow balancing to pump motor.	
10	State the prerequisite for replacement of the mechanical seal of the PCP.	
11	State the preventive maintenance schedule and checks for the pump and motor.	
12	List the prerequisite for replacement of the PHT motor.	
13	List the safety and radiological protection to be taken during overhaul of the pump.	

II.3. Problem solving knowledge

No.	DESCRIPTION	CHECKED BY: (signature and date)
1	State the general causes for increase in thrust bearing temperature of PCP and corrective actions.	
2	Discuss the causes and the checkpoint for the seal leakage of PCPs.	
3	Discuss how to troubleshoot the causes of temperature rise of oil from OSU and their remedies.	
4	State the causal factors for the high seal cavity temperature of the PCP during running and their remedies.	
5	State the reason for low gland flow to the mechanical seal and its remedy.	
6	State the reason for low coast down speed of the PCP and its remedy.	
7	Discuss the probable cause of high temperature in the motor upper thrust and journal bearings and its remedy.	

Appendix VIII

EXAMPLE WORKPLACE ASSESSMENT

Company Specification

Conduct of Training – Workplace Assessment

Originated by: Date:

Reviewed by: Training Standards Working Group Date:

Approved by: Date:
Generation Training Branch Manager

REVISION	AMENDMENT	DATE
000	First issue	June 2001

1. PURPOSE

This training working instruction (TWI) supports integrated company practices (ICPs) BEG/ICP/HR/003 Management of Training Processes and BEG/ICP/HR/004 Deployment of Training Processes.

It supports performance objectives from the ICP, BEG/ICP/HR001 — Performance Objectives and Criteria for the Evaluation and Accreditation of Training Activities — and describes the arrangements required to achieve/support the conduct of workplace assessments.

2. SCOPE

This TWI is applicable to all staff, contractors and organizations associated with the conduct of workplace assessments at all company locations.

3. RESPONSIBILITIES

The Generation Training Manager is responsible for the maintenance of this TWI. The following post holders, groups and teams have specific responsibilities defined in the sections of this TWI:

— Line manager;
— Assessor.

4. PRACTICE

4.1. Trainee assessment

Competence assessment can be used to verify that a person has been able to demonstrate skills, knowledge and behaviour against identified performance criteria. Performance criteria are the skills and behaviour that the candidate must demonstrate.

Performance criteria should be demonstrated under an appropriate range of conditions. An appropriate level of knowledge and understanding are required to demonstrate competence.

4.2. Identification of the need for formal assessment

Consider the job requirements and roles of staff and identify areas where formal competence assessment is required. Where formal training sessions are provided a formal assessment will be required to demonstrate competence.	**Line manager**
Where assessment is appropriate, take advice and guidance from the HR group and agree and review a plan for assessing performance.	**Line manager/ assessor**
Identify an appropriate range of conditions under which the performance criteria should be demonstrated.	**Line manager/ assessor**
Identify the levels of knowledge and understanding required to demonstrate competence.	**Assessor**
Record these details on a competence assessment form.	**Assessor**

4.3. Planning

Ensure that relevant assessment opportunities are identified and selected. As far as possible these opportunities should be carried out in normal work situations, with the minimum disruption to normal work.	**Assessor**
Seek advice from the Training and Development Manager when simulations are unavoidable.	**Assessor**
Ensure that candidates have fair and reliable assessment opportunities.	**Assessor**
Prepare an assessment plan and record the details on the competence assessment form.	**Assessor**
Discuss the assessment plan with the candidate and others who may be affected.	**Assessor**
Review and update the plan to reflect the candidate's progress.	**Assessor**

4.4. Observation

Ensure that all candidates have equal access to assessment opportunities and that they are given the same information.	**Assessor**
Only assess a candidate against the specified performance criteria.	**Assessor**
Assess the candidate accurately against the standards using job instructions, checklists or other reference material where necessary.	**Assessor**
Ensure that any evidence considered is the candidate's own work.	**Assessor**
Ensure that any simulations are correctly set up and carried out.	**Assessor**
Remain as unobtrusive as practicable.	**Assessor**
Give each candidate the same information and judge the evidence fairly and reliably, with no unreasonable discrimination.	**Assessor**
Refer to the HR group in case of difficulties.	**Assessor**
Give clear and constructive feedback on criteria not satisfied and how skills/information may be acquired to ensure success next time.	**Assessor**

4.5. Questioning

Ensure that questions used to demonstrate knowledge are prepared in advance of the observation.	**Assessor**
Identify instances where it is obvious from observing the candidate that he/she can satisfy part of the knowledge requirements.	**Assessor**
Ensure that the questions are valid and address the performance criteria being assessed and that they do not stray into unrelated areas.	**Assessor**
Ensure that questioning is used in conjunction with performance and product evidence to infer competence.	**Assessor**
Ensure preset tests are correctly set up and carried out and that no candidate is placed at a disadvantage.	**Assessor**

Ensure evidence is judged fairly and reliably and clear constructive feedback is given. **Note**: Competence cannot be inferred by questioning alone.	**Assessor**

4.6. Providing feedback

Ensure that the assessment decision is based on the performance and knowledge evidence collected.	**Assessor**
Inform successful candidates when sufficient evidence has been collected to infer competence in all the required situations.	**Assessor**
Provide clear and constructive feedback when candidates are unsuccessful, along with advice on how success may be achieved in the future.	**Assessor**
Ensure that accurate, legible records of evidence are collected and stored in the candidate's personal training record.	**Assessor**

DEFINITIONS

Assessor	A person who has sufficient knowledge and experience to be able to assess the competence of a trainee in performing defined tasks.
Competence	Occurs when an individual can demonstrate a competency. It is the ability to put skills and knowledge into practice in order to perform a job in an effective and efficient manner to an established standard.
Line manager	An individual who has responsibility for the activities of other staff.
Performance training	Optional training.
Permanent record	A permanent record has an identified retention period of 30 years or such other period as approved by the HSE NII and includes records produced in accordance with site licence condition requirements.
Performance objectives and criteria	The performance objectives and criteria to be used for the evaluation and accreditation of training activities carried out under the auspices of the Training Standards and Accreditation Board.
Trainee	A person undertaking a process that will result in a transfer of knowledge or skill.

RELATED PROCEDURES

BEG/ICP/HR/003	Management of Training Processes
BEG/ICP/HR/004	Deployment of Training Processes

RECORDS

Records of formal training will be treated as permanent records.

Appendix IX

EXAMPLE ORAL BOARD INSTRUCTIONS

This is an example of oral board instructions to the trainee. The oral board chair should read these instructions to the trainee at the beginning of the board or provide it to the trainee in advance.

— We are here for your (position) oral board. I am (name) and I will be the board chair. (name) is representing the (organization) and (name) is representing the (other required organizations) . (Introduce any observers that may be present.)
— The examination will take about _____ hours to complete.
— You have been recommended for qualification as a(n) _____ . We are here to determine if you have the necessary knowledge and understanding for that position.
— The scope of this examination is broad in nature and is intended to be the final check of your ability to perform as a qualified operator.
— You may ask for clarification of a question whenever you feel it is necessary.
— Unless otherwise stated by the examiner, assume that the facility is in an operational status with all systems functioning as designed.
— You are expected to answer each question as if you were performing your assigned duties with all available resources and indications. When required, these will be provided for you.
— As you answer each question, you should take some time to arrange your thoughts and answer the question in as much detail as possible. The board members will stop you when they are satisfied with your responses.
— If during your discussion, you should expect to hear an alarm, make or receive a report, or observe some indication, state the fact to us. We will inform you of the indication, report, or alarm status when appropriate.
— If you are unclear on how to answer a particular question, you may pass on that question (i.e. ask to come back to that question later in the board).
— Breaks may be taken whenever you feel it necessary. However, any question that was not answered at the time asked should be answered prior to the break. Unanswered questions will be graded as a failure/ unsatisfactory once a break is taken.
— We will not indicate whether your responses are correct or incorrect.

— We will record all questions and answers. The fact that we are taking notes has no bearing on the correctness of your answer.
— At the completion of the board you will be excused from the room while the board evaluates your performance and determines if you passed or failed.
— Do you have any questions?

Begin the examination.

Appendix X

METHOD FOR ORAL BOARD GRADING

Table 11 provides an example of a method for grading oral boards. It includes criteria for use with a 4.0 scale, a 100% scale and for pass/fail.

Each board member calculates the trainee's area grades using the following method:

TABLE 11. EXAMPLE BOARD GRADING METHOD

Scoring methods			Performance category	Performance description
4.0 scale	% scale	Pass/fail		
<3.2	<80	Fail	Unsatisfactory	The trainee's knowledge of applied fundamental principles, systems, or watch station operation is not acceptable.
≥3.2–<3.5	80–87.5	Pass	Satisfactory/good	A satisfactory/good understanding of the applied fundamental principles, systems, or watch station operation. Some weaknesses.
≥3.5–<3.8	87.6–94.9	Pass	Excellent	Above average understanding of the applied fundamental principles, systems, or watch station operation. A few weak points.
≥3.8–4.0	95–100	Pass	Outstanding	Exceptionally detailed understanding of the applied fundamental principles, systems, or watch station operation. No weak points noted in the trainee's responses.

— Grade each question using a relative importance scale (i.e. questions covering safety systems, technical safety requirements and integrated system operation are of greater importance than questions concerning auxiliary component/system operation).
— To arrive at the area grade, count the grade assigned to the higher importance questions twice and those of lower importance once. Therefore, if two questions were of high importance and five questions were of low importance the total number of grades averaged would be nine. Using this method gives added weight to the high importance questions.
— An example of the grading method mentioned, using HI for high importance questions and LO for low importance questions, is:
  • The trainee received the following grades from a board member; (1) 3.7 (HI); (2) 3.4 (LO); (3) 2.8 (HI); (4) 3.8 (LO); (5) 3.7 (LO); (6) 2.9 (LO); (7) 3.6 (HI); (8) 3.2 (LO).
  • Calculate the area grade average the grades in this manner: 3.7+3.7+3.4+2.8+2.8+3.8+3.7+2.9+3.6+3.6+3.2 = 37.2.
  • Divide 37.2 by 11 and the area grade is 3.38.

Table 12 is an example of an individual board member grading sheet (the above grading method was used to calculate the area grade for the theory and abnormal operations areas which have high and low importance questions). The area grades are given to the board chairperson to determine the overall board grade.

The oral board chairperson should transfer each board member's area grades to the oral board grading sheet and calculate both the overall area grades and the final oral board grade. The oral examination procedure should describe how to derive the overall area grade and the final grade; however, the overall area grade is usually a numerical average of the grades given by each board member in that area. Table 13 is an example of a final board grading sheet. The final board grade is determined in this manner:

— Pass: If all overall area grades are above the minimum necessary to pass.
— Fail: If one or more overall area grades are below the minimum necessary to pass the board. If any overall area grade is less than the minimum required to pass the board, the lowest area grade becomes the grade for the board. For example, if 3.2 is the minimum required to pass and the overall area grades are 3.3, 3.2 and <3.2 the oral board grade would be <3.2 and be classified as a failure.

TABLE 12. INDIVIDUAL BOARD MEMBER GRADING SHEET

Questions by area	Question grade	Area grade
THEORY AREA		
Question 1: Relative importance is high	3.4	3.1
Question 2: Relative importance is high	2.8	
(Two questions were asked, both of high importance, and one was given a failing grade. A failing area grade of 3.1 is calculated.)		
SYSTEM AND COMPONENTS AREA		
Question 1	3.3	3.3
Question 2	3.4	
NORMAL OPERATIONS AREA		
Question 1	3.6	3.4
Question 2	3.3	
ABNORMAL OPERATIONS AREA		
Question 1: Relative importance is high	3.7	3.4
Question 2: Relative importance is low	3.5	
Question 3: Relative importance is low	2.7	
(One high importance and two low importance questions were asked, with one low importance graded as a fail. A passing area grade of 3.4 is calculated.)		
ADMINISTRATIVE CONTROLS AREA		
Question 1	3.3	3.3
Question 2	3.4	
TECHNICAL SAFETY REQUIREMENTS AREA		
Question 1	3.5	3.5
Question 2	3.5	

Print name _____

Signature _____

The board member's signature on the grade sheet signifies that the member participated in the examination of the trainee and that in their opinion the trainee performed as indicated on the grade sheet. At this time the board chairperson should review the grading of each area. If there is significant grading variation (such as one assessor giving a 3.6 and another a 3.0) on the same area, the board chairperson should discuss this variation in grades with

TABLE 13. FINAL BOARD GRADING SHEET

Area	Board member grade	Board member grade	Board member grade	Board member grade (regulatory member)	Board member grade	Overall area grade
Theory	3.1	3.0	3.2	N/A	3.1	3.1
Systems and components	3.3	3.3	3.4	N/A	3.4	3.35
Normal operations	3.4	3.3	3.5	N/A	3.5	3.43
Abnormal operations	3.4	3.4	3.6	N/A	3.4	3.45
Administrative controls	3.3	3.3	3.5	N/A	3.4	3.38
Technical specifications	3.5	3.3	3.2	3.4	3.5	3.38
FINAL BOARD GRADE (see Section 5.5.5 for discussion of grading)						3.1*

* Minimum grade for passing the board is 3.2.

I __(print name)__, Board Chairperson of the oral board for __(print trainee name)__, assign a grade of __ and declare this Board a PASS or FAILURE (circle one).

__(signature Board Chair)__

the board members. Individual grades assigned by the individual board members should be reviewed to determine the reason for the variation. Significant differences should be resolved before an overall board grade is assigned. When differences have been resolved, the grades should be changed on the individual board member grading sheet to reflect the revised grade. Notation on the board grading sheet should be made to explain the correction.

Appendix XI

PERFORMANCE TEST CONSTRUCTION CHECKLIST

Purpose of the test:

1. Does the test require a skill level that is appropriate?

2. Is the purpose of the test clearly stated?

3. Are the objectives of the test clearly stated?

4. Does the test have a clear relationship with the trainee's job or task duties?

Administrative guidelines:

1. Does the test include a segment to help orientate the trainee to the requirements of the test?

2. Have the administrative procedures of the test been clearly spelled out?

Scoring and standards:

1. Have the test's scoring criteria been standardized?

2. Does the test clearly define scoring procedures?

3. Can the scoring rules be quickly applied by the examiner?

4. Have the grading criteria been made as objective as possible?

Instruction section:

1. Does the test contain a complete set of instructions for the trainee?

2. Do the instructions address what is expected of the trainee for perform, simulate and discuss items?

3. Do the instructors address task performance sequence, critical steps and results of failure to comply with safety precautions?

Tools and equipment:

1. Have all required equipment and materials for the test been listed?

2. Has their use been specified at the appropriate level of detail?

Performance steps:

1. Does the test provide a complete and clear listing of all the steps required to perform a task?

2. Are critical and sequential steps identified?

Conditions and cues:

1. Have all initiating and terminating cues been described in the test?

2. Have all environmental conditions been described in the test?

3. Have all equipment conditions been described in the test?

Prerequisites:

Does the test clearly state prerequisite knowledge, experience and skills of the trainee?

Appendix XII

EXAMPLE PERFORMANCE TEST FOR A MAINTENANCE INSPECTION TASK

TEST ID#_____	REV.#_____	PAGE 1 OF 5
TASK #:	TASK TITLE:	AVG. TIME TO PERFORM:
8123506447-3	Inspect lube oil cooler (water side) emergency diesel engine	3 hours

PREREQUISITES:
1. PMT General Course
2. Facility General Course

REFERENCES:
1. NMPT 801.A1 - Emergency Diesel Engine Technical Specification
2. NMPT 801.B: Emergency Diesel Preventive Maintenance Procedure
3. NMPT Form 801: Preventive Maintenance Report
4. NMPT Facility Plan A-804.402
5. NMPT 101.90: System Depressurization

APPROVAL:
Title Date

INSTRUCTIONS TO EVALUATOR:

1. This evaluation standard contains the details for the evaluation of knowledge and practical requirements.
2. Prior to administering the performance test, ensure that the trainee has completed all related training.
3. Prior to administering the performance test, read the Instructions to Examinee to the trainee and review the conditions with him/her. Also, ensure the materials listed are on hand and available for use.
4. When evaluating knowledge requirements, only ask those questions listed in the Knowledge Requirements section. Compare the trainee's response with the answer provided to determine if the question is answered correctly. Failure of any two questions constitutes failure of the performance test.
5. When evaluating task performance, the trainee is expected to perform the steps in sequence. A 'Sat' (satisfactory) can only be awarded after the trainee achieves the listed standard. If a question is included with a step, it must be asked and answered before proceeding with the evaluation. Mark each step as Sat or 'Unsat' (unsatisfactory). Performing any step out of sequence, failing two or more non-critical steps, or any single critical step constitutes a failure of the performance test.
6. Coaching is not permitted during an evaluation. Stop the evaluation when actions may result in damage to personnel or to equipment.
7. Ensure all the information in the scoring/remarks section is correct and complete and that you and the trainee have signed the performance test.
8. Notify the trainee of his/her score immediately upon completion of the performance test.

TEST ID#_____	REV.#_____	PAGE 2 OF 5
TASK #:	TASK TITLE:	AVG. TIME TO PERFORM:
8123506447-3	Inspect lube oil cooler (water side) emergency diesel engine	3 hours

INSTRUCTIONS TO EXAMINEE:

1. The purpose of this performance test is to evaluate your ability to inspect the lube oil cooler (water side) on the emergency diesel engine. The knowledge and skills that are evaluated by this test are directly related to those that you will perform on the job.
2. Before starting, I will state the performance terminal objective and any initiating cues and I will answer any questions you may have.
3. When I tell you to begin, you are to inspect the lube oil cooler (water side) on the emergency diesel engine. I will describe the initial conditions associated with the task and I will ensure that necessary equipment or resources will be available to you.
4. For each step in the task, you are to state what you will do and what result you expect to see in response to your action.
5. At any point, I may stop you and ask you questions regarding the steps, sequence, acceptance criteria, or the effects your actions will have upon the system or component with which you are working or related systems and components.
6. If you perform any two non-critical steps or one critical step improperly or perform a step out of sequence, you will fail this test. If you fail, additional training will be provided and you will be evaluated at a later date.

PERSONNEL/EQUIPMENT SAFETY:
Burn hazard
Damage to diesel
Danger from fluids under pressure

TOOLS/EQUIPMENT:
Protective gloves, face shield, hand tools rigging equipment, lint free rags, crow bar, gasket scraper and gasket material.

PERFORMANCE TERMINAL OBJECTIVE:
Given that applicable references, equipment and materials are available, inspect the lube oil cooler on the emergency diesel engine starting engine in accordance with references 1 through 5.

INITIAL CONDITIONS:
Diesel tagged out and affected portion of lube oil system drained and depressurized.

INITIATING CUES:
Directed by evaluator to begin.

TEST ID#_____	REV.#_____	PAGE 3 OF 5
TASK #:	TASK TITLE:	AVG. TIME TO PERFORM:
8123506447-3	Inspect lube oil cooler (water side) emergency diesel engine	3 hours

PERFORMANCE CHECKLIST

	Action step	Standards	*	Sat	Unsat
P	1. Obtained reference	Trainee obtained latest revised copies of the procedures			
P	2. Obtain equipment	In accordance with NMPT A - 804.402			
P	3. Verify diesel tagged out	In accordance with NMPT 101.90			
P	4. Verify lube oil and cooling water system drained and depressurized	In accordance with NMPT 801.B			
P	5. Disconnect oil supply and return from lube oil cooler	In accordance with NMPT 801.B			
P	6. Disconnect water side supply and return from lube oil cooler	In accordance with NMPT 801.B			
P	7. Remove lube oil cooler from mounting brackets	In accordance with NMPT 801.B			
P	8. Remove lube oil cooler end bells	In accordance with NMPT 801.B			
P	9. Remove lube oil cooler tube bundle	In accordance with NMPT 801.B			
P	10. Perform inspection and document results	In accordance with NMPT Form 801			
P	11. Complete preventive maintenance report	In accordance with NMPT 801.B			
P	12. Reinstall tube bundle	In accordance with NMPT 801.B			
P	13. Reinstall end bell	In accordance with NMPT 801.B			
P	14. Reinstall lube oil cooler into mounting bracket	In accordance with NMPT 801.B			
P	15. Reconnect oil supply and return from lube oil cooler	In accordance with NMPT 801.B			
P	16. Reconnect water side supply and return from lube oil cooler	In accordance with NMPT 801.B			
P	17. Return equipment				
P	18. Return references				

CODES: * (S) Sequence is important. This step must be performed only after the preceding step(s).
(C) Critical step. Failure to meet standards for this item constitutes failure of the test.
P, S, D and O refer to performance methods perform, simulate, discuss and observe/ discuss.

TEST ID#_____	REV.#_____	PAGE 4 OF 5
TASK #:	TASK TITLE:	AVG. TIME TO PERFORM:
8123506447-3	Inspect lube oil cooler (water side) emergency diesel engine	3 hours

KNOWLEDGE REQUIREMENTS:
Directions to Evaluator:
Ask the question exactly as it appears on this form. If the trainee answers differently than the accepted answer, write the exact words of the answer in the comment section.

ORAL QUESTIONS	CRITICAL CONTENT FOR ACCEPTABLE ANSWERS	INITIALS
1. State the precautions associated with removal of the tube bundle. Pass ___ Fail ___ Comments:	1. Tubes are sensitive to impact damage. 2. Damaged tubes may increase oil temperature which may cause damage to diesel.	
Pass ___ Fail ___ Comments:		
Pass ___ Fail ___ Comments:		
Pass ___ Fail ___ Comments:		

TEST ID#_____	REV.#_____	PAGE 5 OF 5
TASK #:	TASK TITLE:	AVG. TIME TO PERFORM:
8123506447-3	Inspect lube oil cooler (water side) emergency diesel engine	3 hours

KNOWLEDGE REQUIREMENTS:

Total number of questions:

Total number of correct responses:

Comments:

PERFORMANCE REQUIREMENTS:

Number of critical steps missed:

Number of non-critical steps missed:

Comments:

OVERALL EVALUATION:

Pass ___ Fail ___

OVERALL COMMENTS:

SIGNATURES:

Evaluators's signature_____ Date _____

Evaluator's name_____ Title_____

Trainee's signature_____ Date _____

Trainee's name _____ ID _____

Appendix XIII*

EXAMPLE PERFORMANCE TEST FOR A SYSTEMS ENGINEER

TECHNICAL DIVISION TRAINING PROGRAMME
PRACTICAL EVALUATION FORM
FOR SYSTEM ENGINEERS
PEV- DSP – 01
ADMINISTRATIVE DUTIES

DUTY AREA: (BSI's) SYSTEM ENGINEER POSITION #

MENTEE NAME _____ **MENTEE SIGNATURE** _____ **EMPLOYEE #** _____	**EVALUATOR NAME AND SIGNATURE** _____ **DATE COMPLETED** _____ (DD MM YYYY)	**PASS** ☐ **FAIL** ☐

Note: When complete, retain a copy of this PEV in the mentee's training file. Upon successful completion of this PEV the evaluator must sign in the box provided and send the original of this cover page to DPAP records office (see SI-TR-21 for instructions)

If there are feedback comments below, forward a copy of this page to the Technical Manager for action.

* Source: Nuclear Power Corporation of India Limited.

Feedback comments (mentee, trainer, evaluator):

	NAME	FUNCTION	SIGNATURE	DATE
PREPARED BY:		**TECHNICAL DELEGATE/ SUPERVISOR**		
APPROVED BY:		**TECHNICAL SUPERINTENDENT**		

PRACTICAL EVALUATION TITLE:	**DOCUMENT #**	**REV:**
System Engineer – Administrative Duties	PEV-DSP-01	0
DIRECTIONS/CONDITIONS	**TIME to PERFORM:** 4 hours	
PREREQUISITE TRAINING: All AB courses identified in the applicable qualification guide passed, credited or exempted.		
PROCEDURES AND REFERENCES REQUIRED: Access to all relevant station documentation given in the objectives of this PEV.		

DIRECTIONS TO EVALUATOR:

The mentee shall demonstrate the following work practice standards during evaluation, as applicable:

- Procedural adherence
- Verbal communication
- Self-checking
- Conventional safety
- Radiation safety
- Conservative decision making
- Teamwork
- Foreign material exclusion
- Housekeeping

Conditions:
This evaluation verifies that the mentee has assimilated the required standard of knowledge and has met the objectives assigned for the subject areas (insert as applicable TECHNICAL/ SPECIALIST KNOWLEDGE, ADMINISTRATIVE DUTIES or SYSTEM SURVEILLANCE) (see IDP-TU-14 for details). The evaluators should be the applicable Technical Supervisor and System Engineer/ Mentor; the Technical Superintendent is responsible for the quality of the evaluation and must sign to confirm that the evaluation is complete.

The evaluation should be based on the objectives assigned in the applicable mentor guides and should be conducted in the office and the field as appropriate. **The mentee will be allowed access to any and all documentation/computer programs that would normally be used in the mentee's day-to-day function.**

As much of the evaluation as possible should be conducted in the English language.

During evaluation sessions, the mentee must demonstrate a high standard of work practice expectations as applicable, especially in the areas of quality, safety and responsibility.

During evaluation sessions, the mentee should not be prompted or coached but must be given sufficient time to complete the evaluation tasks. The mentee is required to perform any tasks using approved procedures.

During field tours the evaluator should ensure that the mentee is aware of and uses safety equipment and processes as appropriate — the safety of the mentee is the responsibility of the evaluator.

TOOLS AND EQUIPMENT REQUIRED:

As listed in the quoted reference documents, all necessary documents are to be provided to the mentee.

PRACTICAL EVALUATION TITLE: System Engineer – Administrative Duties	DOCUMENT # PEV-DSP-01	REV: 0
STANDARDS	TIME to PERFORM: 4 hours	

STANDARDS:
As per the standards given in each of the objectives.

COMMON STANDARD FOR RESPONSE TO QUESTIONS:

The mentee must demonstrate a satisfactory knowledge and understanding of the following:

— Underlying principles associated with the duty area/task being evaluated;

— Application of these principles to competent performance in the duty area/task being evaluated;

— Procedures and standards governing task performance.

The following are typical examples of the types of question that the examiner may ask during the practical evaluation.
(Reference indicates a procedure, course or source where the answer to the question may be obtained, if applicable).

EVALUATOR DEBRIEFINGS:

Evaluator debriefing mentee on practical evaluation:

- Evaluator informs mentee whether evaluation is successful or unsuccessful;
- Evaluator should make positive comments on evaluation performance;
- Evaluator should suggest improvements as appropriate;
- Evaluator signs approved section of Practical Evaluation form and indicates PASS/FAIL.

If the result of this evaluation is a failure then the evaluator must give the mentee detailed instructions on how to improve performance and when the next evaluation will take place.

PRACTICAL EVALUATION TITLE: System Engineer – Administrative Duties	**DOCUMENT #** PEV-DSP-01	**REV:** 0
OBJECTIVES	**TIME to PERFORM:** 4 hours	

ACTIONS/QUESTIONS	**STANDARDS**	**CRITICAL**	**PASS**	**FAIL**
1. After working as a system engineer in training for a period of about 3 to 6 months, the mentee will describe the organization and its roles and responsibilities to the evaluator.	The candidate will: — Give feedback on the data given on the TD General Information Guide. — Describe satisfactorily how he/she has fulfilled the roles and responsibilities for the assigned systems. — Describe the actual interfacing carried out with: • Maintenance • Operations • Planning • Design engineering • Safety and compliance • Procurement.			
2. Discuss each of the following items as they relate to performing the system engineer function: — IDP-TU-14 training and qualification programme for technical division staff. Procedures/governing documents. — Planning, work control and outage planning. — Assisting operations.	Common standard for all (previous page), plus demonstrates an understanding of the following references and procedures: — In accordance with the applicable document listing in Mentor Guide MG-DSP-OOO. — In accordance with the applicable document listing in Mentor Guide MG-DSP-003. — In accordance with the applicable document listing in Mentor Guide MG-DSP-004.			

PRACTICAL EVALUATION TITLE: System Engineer – Administrative Duties	**DOCUMENT #** PEV-DSP-01	**REV:** 0
OBJECTIVES	**TIME to PERFORM:** 4 hours	

ACTIONS/QUESTIONS	**STANDARDS**	**CRITICAL**	**PASS**	**FAIL**
Assisting maintenance	In accordance with the applicable document listing in Mentor Guide MG-DSP-002.			
Technical operability evaluations	In accordance with the applicable document listing in Mentor Guide MG-DSP-011.			
Configuration management	In accordance with the applicable document listing in Mentor Guide MG-DSP-008.			
Modifications	In accordance with the applicable document listing in Mentor Guide MG-DSP-006.			
Corrective actions programme	In accordance with the applicable document listing in Mentor Guide MG-DSP-000.			
Records management	In accordance with the applicable document listing in Mentor Guide MG-DSP-000.			
Nuclear safety and compliance	In accordance with the applicable document listing in Mentor Guide MG-DSP-000.			

Appendix XIV*

EXAMPLE INSTRUCTOR OBSERVATION FORM

Instructor performance evaluation

LESSON TITLE: _____ DATE: _____

INSTRUCTOR: _____ LENGTH OF OBSERVATION:_____

OBSERVED BY: _____ DATE: _____

REVIEWED BY: _____ DATE: _____

Instructions: Below is a list of competency statements that instructors should use to contribute to the learning process. Read each statement and evaluate the instructor's performance by circling the appropriate rating next to the statement. Written comments for all ratings are encouraged. Comments are required for 'unsatisfactory' and 'needs improvement' ratings. Space is available to the right of each rating.

EXPLANATION OF RATINGS

0 Not observed: Activity not observed by the evaluator
1 Unsatisfactory: Failed to perform the required activity
2 Needs improvement: Performed most essential activities properly
3 Satisfactory: Performed all essential activities properly
4 Above average: Performed all requirements and exceeds on several
5 Outstanding: Consistently exceeded requirements

MATERIALS	COMMENTS
1. The student handout is organized in a logical manner conforming with lesson presentation.	0 1 2 3 4 5
2. The training material is current and technically accurate.	0 1 2 3 4 5
3. The training material relates to the learning objectives.	0 1 2 3 4 5
4. When used, the industry event examples are appropriate.	0 1 2 3 4 5

* Source: Électricité de France.

CONDUCT OF CLASS

Preparation

1.	Classroom physical layout enhanced the learning climate.	0 1 2 3 4 5
2.	The instructor appeared adequately prepared.	0 1 2 3 4 5

Introduction

1.	Started class on time.	0 1 2 3 4 5
2.	Provided student handouts.	0 1 2 3 4 5
3.	Stated the purpose of the lecture.	0 1 2 3 4 5
4.	Reviewed the objectives for the class session.	0 1 2 3 4 5
5.	Stated a problem to be solved or discussed during the class.	0 1 2 3 4 5
6.	Made explicit the relationship between current subject matter and previous classes.	0 1 2 3 4 5

Presentation

1.	Followed the lesson plan.	0 1 2 3 4 5
2.	Taught the content in a systematic and organized fashion.	0 1 2 3 4 5
3.	Defined new terms, concepts and principles.	0 1 2 3 4 5
4.	Used clear, simple and relevant examples to explain major ideas.	0 1 2 3 4 5
5.	Related new ideas to familiar ones.	0 1 2 3 4 5
6.	Presented information at an appropriate level of detail.	0 1 2 3 4 5
7.	Used alternative explanations when necessary.	0 1 2 3 4 5
8.	Stated the relationship between various ideas in the presentation.	0 1 2 3 4 5
9.	Asked questions to determine if information was presented at a proper rate.	0 1 2 3 4 5
10.	Periodically summarized the important ideas.	0 1 2 3 4 5

11.	Reiterated definitions of new terms to help students become accustomed to them.	0 1 2 3 4 5
12.	Exhibited a level of knowledge adequate to teach the material.	0 1 2 3 4 5
13.	Displayed a positive attitude.	0 1 2 3 4 5
14.	Demonstrated confidence during the class presentation.	0 1 2 3 4 5
15.	Developed a positive rapport with the students.	0 1 2 3 4 5
16.	Encouraged student participation.	0 1 2 3 4 5

COMMUNICATION SKILLS

Verbal

1.	Voice could be easily heard.	0 1 2 3 4 5
2.	Voice was raised or lowered for variety and emphasis.	0 1 2 3 4 5
3.	Speech was neither too formal nor too casual.	0 1 2 3 4 5
4.	Rate of speech was neither too fast nor too slow.	0 1 2 3 4 5
5.	Varied the pace of the presentation to keep the students alert.	0 1 2 3 4 5
6.	Spoke at a rate that allowed students time to take notes.	0 1 2 3 4 5
7.	Facilitated discussions effectively.	0 1 2 3 4 5

Non-verbal

1.	Established and maintained eye contact with the entire class.	0 1 2 3 4 5
2.	Listened carefully to student comments and questions.	0 1 2 3 4 5
3.	Appearance was proper.	0 1 2 3 4 5
4.	Instructor was enthusiastic about the material presented.	0 1 2 3 4 5
5.	Noted and responded to signs of puzzlement, boredom and curiosity among the students.	0 1 2 3 4 5

QUESTIONING ABILITY

Asking questions

1.	Asked questions to determine what the students know about the lecture topic.	0 1 2 3 4 5
2.	Asked questions that allowed the instructor to gauge student progress.	0 1 2 3 4 5
3.	Addressed questions to individual students as well as to the group at large.	0 1 2 3 4 5
4.	Encouraged students to answer difficult questions by providing clues or rephrasing.	0 1 2 3 4 5
5.	Involved as many students as possible in the classroom discussion.	0 1 2 3 4 5
6.	When necessary, asked students to clarify their questions.	0 1 2 3 4 5
7.	Asked probing questions if a student's answer was incomplete or superficial.	0 1 2 3 4 5
8.	Repeated answers when necessary so the entire class could hear.	0 1 2 3 4 5

Answering questions

1.	Encouraged student questions.	0 1 2 3 4 5
2.	Received student questions politely and, when possible, enthusiastically.	0 1 2 3 4 5
3.	Answered student's questions satisfactorily.	0 1 2 3 4 5
4.	Repeated student's question when necessary.	0 1 2 3 4 5

AUDIO/VISUAL AIDS

1.	Used audio/visual aids to enhance the learning objectives.	0 1 2 3 4 5
2.	Transparencies/slides were clear and easy to read.	0 1 2 3 4 5
3.	Board work appeared organized and legible.	0 1 2 3 4 5
4.	Demonstration performed could be seen by all students.	0 1 2 3 4 5
5.	Student handouts were used effectively by the instructor.	0 1 2 3 4 5

SUMMARY

1. The instructor properly summarized the key points of the presentation. 0 1 2 3 4 5

STRONG POINTS	SUGGESTIONS FOR IMPROVEMENT

Overall Evaluation Score

1	2	3	4	5
Needs improvement		Minimum requirement		Exceeded requirement

Signature of trainee Date

Signature of evaluator Date

Appendix XV*

SELF-ASSESSMENT OF INSTRUCTOR QUALITIES

Take your time and think about each of the statements listed below. Then determine how your actions compare with those described. Indicate your score for each of the statements by selecting the number that most closely corresponds to you. Be honest with yourself. No one else will have to see your answers and very often no two answers are the same for each statement. This exercise is meant to give you an idea of how your toolboxes of instructor skills compare with others who have been successful as instructors. Compare your total score to those in the box at the bottom.

		Always		Some-times		Never	
1.	I am fair, consistent and non-judgemental in my dealings with trainees.	5	4	3	2	1	0
2.	I believe in my trainees and encourage them whenever possible.	5	4	3	2	1	0
3.	I display enthusiasm for the subject matter and show that I am pleased when the trainees learn the material.	5	4	3	2	1	0
4.	I am knowledgeable in my subject matter and remain current with modifications, industry experience and changes in the job requirements.	5	4	3	2	1	0
5.	I am prepared and organized whenever I present material to the trainees.	5	4	3	2	1	0
6.	I allow the trainees to get involved in the planning and choosing of activities whenever possible.	5	4	3	2	1	0
7.	I look for ways to improve my ability to prepare and present instructional material that is clear, organized and important to job performance.	5	4	3	2	1	0
8.	I make my instruction relevant to the real life needs of my trainees and the company.	5	4	3	2	1	0
9.	My classroom is a comfortable and challenging experience, but free from unnecessary pressure.	5	4	3	2	1	0

* Source: USDOE.

10.	I always maintain and model a professional appearance and attitude.	5	4	3	2	1	0
11.	I strive to remain current with the latest techniques for effective training.	5	4	3	2	1	0
12.	I continuously try to improve my speaking and presentation skills.	5	4	3	2	1	0
13.	I have patience when dealing with trainee's problems and display empathy for the trainee point of view.	5	4	3	2	1	0
14.	I strive to make my training meaningful, relevant and enjoyable.	5	4	3	2	1	0
15.	I make myself available to answer questions and address concerns of the trainees without making them feel as if they are infringing on my time.	5	4	3	2	1	0
	Column totals:						

Total score _____

65–75	You possess many of the qualities that help you to be a successful instructor. Build on your qualities and continue to work on improving your skills.
55–64	You should identify the areas where you need to improve your qualities. Any effort will quickly help you improve as an instructor.
45–54	You should formulate a specific action plan to develop your weak areas. You need to develop a broad base of qualities to help you succeed.
<45	You should recognize the need to develop these qualities. Begin now by identifying several areas you will concentrate on right away. Even though these qualities will not guarantee your success, they can help you in your training efforts.

REFERENCES

[1] INTERNATIONAL ATOMIC ENERGY AGENCY, Recruitment, Qualification and Training of Personnel for Nuclear Power Plants, IAEA Safety Standards Series No. NS-G-2.8, IAEA, Vienna (2002).

[2] INTERNATIONAL ATOMIC ENERGY AGENCY, Nuclear Power Plant Personnel Training and its Evaluation: A Guidebook, Technical Reports Series No. 380, IAEA, Vienna (1996).

[3] INTERNATIONAL ATOMIC ENERGY AGENCY, Training the Staff of the Regulatory Body for Nuclear Facilities: A Competency Framework, IAEA-TECDOC-1254, IAEA, Vienna (2001).

[4] INTERNATIONAL ATOMIC ENERGY AGENCY, Development of Instructors for Nuclear Power Plant Personnel Training, IAEA-TECDOC-1392, IAEA, Vienna (2004).

[5] INTERNATIONAL ATOMIC ENERGY AGENCY, Assuring the Competence of Nuclear Power Plant Contractor Personnel, IAEA-TECDOC-1232, IAEA, Vienna (2001).

[6] INTERNATIONAL ATOMIC ENERGY AGENCY, Selection, Competency Development and Assessment of Nuclear Power Plant Managers, IAEA-TECDOC-1024, IAEA, Vienna (1998).

[7] INTERNATIONAL ATOMIC ENERGY AGENCY, Use of Control Room Simulators for Training of Nuclear Power Plant Personnel, IAEA-TECDOC-1411, IAEA, Vienna (2004).

[8] THALHEIMER, W., The Learning Benefits of Questions, Work-Learning Research (2003), http://www.work-learning.com/ma/PP_WP003.asp

[9] BLOOM, B.S., ENGLEHART, M.D., FURST, E.J., HILL, W.H., KRATHWOHL, D.R., Taxonomy of Instructional Objectives: Handbook 1: Cognitive Domain, David McKay, New York (1956).

[10] KRATHWOHL, D.R., BLOOM, B.S., MASIA, B.B., Taxonomy of Educational Objectives: Handbook 2: Affective Domain, David McKay, New York (1964).

[11] SIMPSON, E.J., The Classification of Educational Objectives in the Psychomotor Domain, Grypon House, Washington, DC (1972).

[12] ERICKSON, H., MILLER, P., Test Development: Guidelines and Practical Suggestions, Patrick W. Miller and Associates, Munster, IN (2001).

[13] KUBISZYN, T., BORICH, G., Educational Testing and Measurement, 7th edn, John Wiley, New York (2003).

[14] INTERNATIONAL ATOMIC ENERGY AGENCY, Means of Evaluating and Improving the Effectiveness of Training of Nuclear Power Plant Personnel, IAEA-TECDOC-1358, IAEA, Vienna (2003).

BIBLIOGRAPHY

AIRASIAN, P., Classroom Assessment: Concepts and Applications, McGraw-Hill, New York (2004).

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, AMERICAN PSYCHOLOGICAL ASSOCIATION, NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION, 1999 Standards for Educational and Psychological Testing, AERA, Washington, DC (2000).

BAKER, F., The Basics of Item Response Theory (2001), http://www.edres.org/irt/

HALADYNA, T., Developing and Validating Multiple Choice Test Items, 2nd edn, Lawrence Erlbaum Associates, Mahwah, NJ (1999).

Institute of Objective Measurement Inc. (2005), http://www.rasch.org

LANDY, F.J., Psychology of Work Behavior, Dorsey Press, Homewood, IL (1985).

NUCLEAR REGULATORY COMMISSION, Examiners Handbook for Developing Operator Licensing Examinations, Rep. NUREG/BR-0122, Rev. 5, NRC, Washington, DC (1990).

— Operator Licensing Examination Standards for Power Reactors, Rep. NUREG-1021, Rev. 9, NRC, Washington, DC (2003).

PEDHAZUR, E., Measurement Design and Analysis, Lawrence Erlbaum Associates, Mahwah, NJ (1991).

STARK, S., et al., Item Response Theory, http://work.psych.uiuc.edu/irt/

THORNDIKE, R.M., Measurement and Evaluation in Psychology and Education, 7th edn, Macmillian, New York (2004).

UNITED STATES DEPARTMENT OF ENERGY, Guide to Good Practices for the Development of Test Items, Rep. DOE-HDBK-1204-97, Washington, DC (1997).

— Guide to Good Practices for the Design, Development, and Implementation of Examinations, Rep. DOE-HDBK-1205-97, Washington, DC (1997).

— Guide to Good Practices for Oral Examinations, Rep. DOE-HDBK-1080-97, Washington, DC (1997).

UNITED STATES DEPARTMENT OF LABOR, TESTING AND ASSESSMENT, An Employers Guide to Good Practices (1999), http://www.onetcenter.org/dl_files/empTestAsse.pdf

WRIGHT, B.D., STONE, M.H., Best Test Design, Rasch Measurement Series, Mesa Press, Chicago (1979).

CONTENTS OF THE ACCOMPANYING CD-ROM

CODE ELEMENTS FOR COUNTRY NAMES – Abstracted from ISO-3166 Part I (1997)

CA – Canada
CH – China
DE – Germany
ES – Spain
GB – Great Britain
IN – India
PK – Pakistan
RF – Russian Federation
RO – Romania
US – United States of America

DOCUMENTS AND PROCEDURES PROVIDED BY MEMBER STATES

No.	Country	Title	Reference section
1	CA	Operational Procedure CNSC-STI, Rev. 2.2 Written and Oral Examinations for Certified Operating Personnel at Nuclear Reactor Facilities, July 2002 (125 pages)	4, 5
2	CA	CNSC-EG1: for written examination	4
3	CA	Examination Guide CNSC-EG2, Rev. 0 (185 pages)- Requirements and Guidelines for Simulator-Based Certification Examinations for Shift Personnel at Nuclear Power Plants	6
4	CA	CNSC Development Programme for Recruitment of Young Engineers and Scientists	2, 6
5	CA	Computer Based Training Development and the CBT Testing Process at Ontario Power Generation (20 pages)	7
6	CH	Daya Bay regulation of examination and evaluation with example evaluation and assessment sheets	4, 5, 6

DOCUMENTS AND PROCEDURES PROVIDED BY MEMBER STATES (cont.)

No.	Country	Title	Reference section
7	CH	RINPO Staff Training, Examination and Qualification (22 pages in Chinese)	2, 4, 5, 6
8	DE	Guideline on the Content of the Qualification Examinations of responsible Shift Personnel in NPPs 1996	2, 4
9	DE	Extracts from the Catalogue of Questions for the Examinations of Nuclear Fundamentals	4
10	ES	GSG-01.01 Cualificaciones para la obtención y uso de licencias de personal de operación en centrales nucleares (20 pages in English and 7 pages in Spanish) http://www.csn.es/./plantillas/ frame_publicaciones.jsp?id_nodo=280&&tipo=GSG&keyword=&auditoria=F	3, 4, 5, 6
11	GB	British Energy, Conduct of Simulator Assessments, January 31, 2001 (19 pages)	6
12	GB	British Energy, Conduct of Training-Workplace Assessment (6 pages)	6
13	IN	Quality Assurance of Licensing Examinations-Syllabus & Question Papers Part 1, April 1992 (6 pages)	3
14	IN	Competency Verification of Operation Personnel Focusing CE (Control Engineer=Reactor Operator) as case study, presentation by D.V. Vishwaroop- March 14, 2005 (Syllabus for main plant Operation Personnel	2, 3, 4, 5, 6
15	IN	On-Job-Training (OJT) cum checklist programs (4 pages)	5, 6
16	IN	Field Checklist for Operation, Control Engineer (Main Plant) OP-III-7 (17) and OP-III-9 (18) Reactivity Mechanism – Perform manual operation of adjuster rod (5 pages)	5, 6
17	IN	Field Checklist for Operation, Control Engineer (Main Plant) OP-III-9 (18) Reactivity Mechanism – Perform shut off rod drop time test (5 pages)	5, 6
18	IN	Field Checklist for Operation, Control Engineer (Main Plant) OP-III-24 (74) Shutdown system (SDS#2) Process – Perform back flushing of one poison tanks during unit operation (3 pages)	5, 6

DOCUMENTS AND PROCEDURES PROVIDED BY MEMBER STATES (cont.)

No.	Country	Title	Reference section
19	IN	Evaluation of Trainers in Indian NPPs with examination syllabus, position specific task lists, and sample task performance checklist (26 pages)	3, 4, 5, 6
20	IN	Short-listed Task List for trainers – Appendix B (20 pages)	3
21	IN	Nuclear System Surveillance and Event Management (Paper No. OP-II-1)-describes knowledge areas for assessments (6 pages)	5
22	IN	Observation of task performance by trainers (2 pages)	5, 6
23	IN	Checklist for Chemistry Control – Verifying gas concentrations (4 pages)	5, 6
24	IN	Checklist for Radiation Protection – Perform heavy water sample analysis (4 pages)	5, 6
25	IN	Checklist for maintenance – Qualify to perform critical maintenance on PHT system (4 pages)	5, 6
26	PK	Booklet-Professional Training course for PRNA Inspectors (Level 1) covering nuclear safety, radiation safety and regulatory control (4 pages)	3
27	PK	Detailed content of course (13 pages)	3
28	PK	Example Mid-Term Test, 65 questions	4
29	PK	Example Terminal Examination Paper, 100 questions (12 pages)	4
30	PK	Syllabus for Level II Course on primary and auxiliary systems and secondary systems (4 pages)	3
31	PK	Example final paper, 12 questions on primary and auxiliary systems (4 pages)	4
32	PK	Example Final Paper, 19 questions on secondary systems	4
33	RO	Instructions for Preparing Qualification Guides and Practical Evaluation Guides, IDP-TU-21 (9 pages)	6
34	RO	Practical Evaluation Form for System Engineers, PEV-DSP-01, Administrative Duties (5 pages)	6
35	RF	Description of Computer Based Testing System, “Barrier” (24 pages) for top managers and operating personnel	7

DOCUMENTS AND PROCEDURES PROVIDED BY MEMBER STATES (cont.)

No.	Country	Title	Reference section
36	RF	Description of Computer Based testing System, "Vector" for operating personnel (8 pages)	7
37	RF	Test Item Performance (Russian and English) – Example of test item analysis.	10
38	US	U.S. Department of Labor, Testing and Assessment: An Employers Guide to Good Practices http://wdr.doleta.gov/opr/fulltext/document.cfm?docn=6032	2, 3, 4, 9, 10
39	US	U. S. Nuclear Regulatory Commission, NUREG-1021, Operator Licensing Examination Standards for Power Reactors http://www.nrc.gov/reactors/operator-licensing.html (444 pages)	4, 5, 6, 9
40	US	U.S. Nuclear Regulatory Commission, NUREG-1122, Revision 2, knowledge and Abilities Catalog for Nuclear Power Plant Operators: Pressurized Water Reactors (461 pages)	3
41	US	U.S. Nuclear Regulatory Commission Generic Fundamentals Examination Bank—Pressurized Water Reactors	4
42	US	Thalheimer, W., The Learning Benefits of Questions, January 2003, http://www.work-learning.com/ma/PP_WP003.asp	2, 3
43	US	PPL Susquehanna NPP, Nuclear Training Procedure, Examination Administration and Academic Honesty Policy, NTP-QA-14.2 (41 pages)	4, 5, 6
44	US	PPL Susquehanna NPP, Nuclear Training Procedure, Selection Examination Development, Validation, and Analysis, NTP-QA-11.9 (15 pages)	3, 9
45	US	Test Item Formats	4
46	US	Sample Test Specification for a 50-item General Employee Test	4
47	US	Multiple Choice Assessment Question Review Checklist	4
48	US	JPM Trainee Assessment (3 pages)	6
49	US	Exelon example – out of the box evaluation of human performance training (4 pages)	5, 6

CONTRIBUTORS TO DRAFTING AND REVIEW

Bastos, J.L.F.	International Atomic Energy Agency
Dai, Z.	CGNPC, China
Halada, J.	VUJE, Slovakia
Huang, F.	International Atomic Energy Agency
Janezic, A.	Slovenian Nuclear Safety Administration, Slovenia
Kiss, I.	Paks Nuclear Power Plant, Hungary
Kraut, A.	GRS GmbH, Germany
Mazour, T.	International Atomic Energy Agency
Molloy, B.	Consultant, United Kingdom
Petelin, S.	University of Ljubljana, Slovenia
Rodriguez Diaz, J.A.	CSN, Spain
Shaw, D.	CNSC, Canada
Szabo, V.	Nuclear Regulatory Authority, Slovakia
Tiron, C.	Nuclearelectrica S.A., Romania
Vishwaroop, D.	NPCIL, India
Yoder, J.	Consultant, United States of America

Technical Meeting

Vienna, Austria: 14–17 March 2005

The nuclear industry spends a significant amount of resources conducting competency assessments. Competency assessments are used for employee selection, trainee assessment, qualification, requalification and authorization. This publication focuses on competency assessments used for measuring the knowledge, skills and attitudes of personnel as the result of training. Other uses of competency assessments are also briefly discussed. While not intended to provide in-depth coverage of assessment theory, this publication — along with the attached CD-ROM — should provide developers, instructors and assessors with a foundation on which to develop sound assessments.

INTERNATIONAL ATOMIC ENERGY AGENCY
VIENNA
ISBN 92–0–110105–8