

A photograph of the California State Capitol building, showing its iconic white dome with a gold top and classical architectural details like columns and arches. The building is set against a clear blue sky. A dark blue curved graphic element separates the image from the white background of the text.

Testing in Core Course Participant Workbook 2020

Testing in Standards and Training for Corrections Core Courses

Course Overview

Instructors for Standards and Training for Corrections (STC) Core Courses are required to develop tests that are used to assess the outcome of core training. In this course, you will learn general guidelines and principles for developing the three types of tests utilized in core courses. The basic principles of test development, guidelines for development, and key aspects for review and evaluation will be covered. Additionally, guided practice opportunities for developing the three types of test will be provided.

Performance Objectives

Upon completion of this course, you will have the knowledge and skills to:

- describe basic principles of test development.
- identify three types of Multiple Choice Tests that are used to assess core course learning objectives.
- develop multiple choice test items that adhere to the multiple-choice item writing guidelines.
- identify key aspects for reviewing and evaluating test items.

Table of Contents

Types of Tests in STC Core Courses	1
Behavior Skills Test (BST).....	1
Written Skills Test (WST).....	1
Multiple Choice Test (MCT).....	2
Basic Principles of Test Development.....	3
Core Courses as a Selection Standard	3
Legislation and Personnel Selection.....	4
Legal Requirements and Professional Standards for Employee Selection	5
Application to Testing in STC Core Courses	5
Best Practices for Testing.....	7
STC Requirements and Recommendations.....	7
Edit and Review.....	7
Reviewing Test Performance	8
Test Security	9
Structure of Multiple-Choice Items.....	10
Group Exercise 1	11
Multiple-Choice Formats	12
Conventional MC.....	12
Matching.....	13
Extended Matching.....	14
Alternate Choice.....	15
True/False.....	15
Multiple True/False	16
Multiple-Choice Item Writing Guidelines.....	17
Content Guidelines.....	17
Format and Style Guidelines	21
Stem Guidelines	24
Alternative Guidelines	26
Group Exercise 2	33
Group Exercise 3	35
Guidelines for Designing BSTs and WSTs	36
Solicit Appropriate Trainee Response.....	36
Measure the Core Course Performance Objective.....	37

Provide the Resources Specified in the Performance Objective.....	38
Assess Important Content.....	39
Pass the Reasonable Person Test	40
References	41
Additional Resources.....	42

Types of Tests in STC Core Courses

Most performance objectives in the core training manuals are followed by the type of test that is required to assess each student's ability to accomplish the performance objective. This section describes the three types of tests including development responsibility and time of delivery.

Behavior Skills Test (BST)

BSTs require the trainee to physically demonstrate a job skill such as donning a self-contained breathing apparatus. The design of BSTs is guided by or laid out in the performance objectives. For example, consider BST #8 from the Adult Corrections Officer (ACO) Core Course provided below. The instructor would provide the instruction and guided practice necessary for trainees to learn the appropriate techniques to perform a rear break fall. Then each trainee would be responsible for individually demonstrating a rear break fall. The components listed in the bullets shall be demonstrated during the rear break fall. BSTs are completed *during* classroom instruction time and are graded by the instructor as pass/fail.

BST ACO Core Course

ACO Core Performance Objective: In a simulation of a frontal assault, demonstrate a rear break fall technique incorporating the following:

- correct body positions
- proper movement
- balance
- position of advantage

Written Skills Test (WST)

WSTs require the trainee to write a simulated job document such as a portion of a report, interview questions, or an incident report. The design of WSTs is guided by the performance objectives. For example, consider the WST from the ACO Core Course provided below. The instructor would provide a sample case description and policy. Each trainee would individually classify the individual in the description and provide supporting rationale. The instructor would develop the sample case description and sample department classification policy. This WST would be completed during the classroom instruction time and graded by the instructor during non-training hours.

WST ACO Core Course

ACO Core Performance Objective: Given a sample case description, classify the individual using a sample department classification policy and state your rationale.

WSTs and any supporting materials are developed by the instructor who is responsible for the delivery of the associated module. WSTs are completed *during* classroom instruction (time to complete the WST is built into the hourly outline) and graded (pass/fail) by the instructor during non-training hours. It is not acceptable to permit trainees to complete these tests at home because hours outside of classroom work may conflict with Fair Labor Standards Act issues regarding overtime and there is no way to ensure that the trainee is actually the person who completes the test.

Multiple Choice Test (MCT)

MCTs are paper and pencil tests that are usually presented in a multiple-choice format. The multiple-choice test (MCT) format is the preferred methodology for measuring knowledge-based performance objectives because of its proven value in measuring different types of learning, ease of administration, and scoring. An MCT is comprised of many questions, commonly referred to as *items*, that when combined, assess multiple performance objectives from one or more course modules or units.

In the ACO, PO, and JCO Core Course Manuals, performance objectives that are assessed with an MCT requires a minimum of one item to assess the performance objective.

Example

JCO Core Performance Objective: Describe the steps in the Juvenile Justice System from arrest to disposition. MCT 1 item

In most cases, instructors for each core course module should develop the multiple-choice items that are combined to develop the MCTs. Instructors will be able to link the items most accurately to the performance objective and lesson plans. It is also recommended that instructors develop a bank of items to draw on to prepare for remediation, substituting for poorly written items, and alternating items for test security. Each core course has time designated for the administration of MCTs. Thus, these tests are administered after the delivery of the associated content (i.e., module or unit).

Basic Principles of Test Development

Core Courses as a Selection Standard

Role of STC: Penal Code 6035

(a) For the purpose of raising the level of competence of local corrections and probation officers and other correctional personnel, the board shall adopt, and may from time to time amend, rules establishing minimum standards for the selection and training of these personnel employed by any city, county, or city and county who provide for the custody, supervision, treatment, or rehabilitation of persons accused of, or adjudged responsible for, criminal or delinquent conduct who are currently under local jurisdiction. All of these rules shall be adopted and amended pursuant to Chapter 3.5 (commencing with Section 11340) of Part 1 of Division 3 of Title 2 of the Government Code.

California Code of Regulations: Title 15 § 131. Minimum Selection Standards

(a) In addition to the requirements set forth in Section 830 et seq. of the Penal Code and Section 1029 et seq. of the Government Code, the standards set forth below shall apply. These standards for entry probation officer positions, entry juvenile counselor positions, and entry corrections officer positions shall include but not be limited to the following:

- (1) Basic abilities and other characteristics important for successful job performance as demonstrated by passing the Board's written examination. An alternative examination may be substituted pursuant to Section 132.
- (2) Competence in oral communication as demonstrated in an interview.
- (3) Past behavior compatible to job requirements as demonstrated by a background investigation.
- (4) Competence in the knowledge, skills and abilities necessary for entry-level job performance, as demonstrated by successful completion of the required core curriculum.
- (5) Competence in the performance of entry-level duties as demonstrated by successful completion of the probationary period.
- (6) The ability to perform the essential job functions of the position as demonstrated by meeting the Board's current guidelines for Vision, Hearing, and Medical Screening.
- (7) A minimum of 18 years of age prior to appointment.

(b) The level of competence in (a) (2) and (a) (3) above shall be commensurate with the needs of the individual job classifications of each county or city. The level of competence in (a) (1) above shall be commensurate to the cutoff score that is chosen by the county or city and is consistent with research validation.

Legislation and Personnel Selection

Fair Employment Law

1964 Civil Rights Act – had many titles, the most important for personnel decisions was Title VII, which:

- made it illegal to discriminate based on race, color, religion, sex, or national origin as a basis for making hiring decisions.
- created protected groups and subgroups including Blacks, American Indians (including Alaskan Natives), Asians (including Pacific Islanders), Hispanic (including persons of Mexican, Puerto Rican, Cuban, Central or South American, or other Spanish origin or culture regardless of race), and whites (Caucasian) other than Hispanic.
- created the Equal Employment Opportunity Commission (EEOC) for oversight and enforcement.
- defined selection procedures.

Selection Procedure

A selection procedure is any measure, combination of measures, or procedure used as a basis for any employment decision. Selection procedures include the full range of assessment techniques from traditional paper and pencil tests, performance tests, **training programs**, or probationary periods and physical, educational, and work experience requirements through informal or casual interviews and unscored application forms.

As a result of Title VII, there was and continues to be a close examination of the procedures that are used for making hiring decisions. This examination has mostly happened through litigation settled at the Supreme Court Level. Additional protected groups were identified as a result of the:

- 1967 Age Discrimination in Employment Act – This forbids discrimination against people who are age 40 or over.
- 1990 Americans with Disabilities Act – This prohibits discrimination against people who have disabilities. “A disabled person is defined as one with a physical or mental impairment that substantially limits one or more major life activities, or who has a record of such impairment, or who is regarded as having such impairment (Guion, 1998, p. 199).” This does not protect the employment of those who would threaten the safety or property of others on a given job.

Outcomes of Important U.S. Court Cases

- Procedures must be job related as demonstrated through research.
- Defendants are responsible for providing validity evidence.
- Procedures may not wrongfully discriminate.
- Definition of a test – A test is **any form** of collecting information on individuals when that information is used as a basis for making an employment decision.

Legal Requirements and Professional Standards for Employee Selection

Uniform Guidelines on Employee Selection Procedures [Equal Employment Opportunity Commission (EEOC), Civil Service Commission (CSC), Department of Labor (DOL), & Department of Justice (DOJ), 1978]

- These provide a single set of principles which are designed to assist employers, labor organizations, employment agencies, and licensing and certification boards to comply with requirements of Federal law, prohibiting employment practices which discriminate on the grounds of race, color, religion, sex, and national origin.
- They were designed to provide a framework for determining the proper use of tests and other selection procedures.
- They are consistent with professional standards, such as those described in the Standards for Educational and Psychological Testing and the Principles for the Validation and Use of Personnel Selection Procedures.

Standards for Educational and Psychological Testing [American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME), 1999]

- These provide criteria for the evaluation of tests, testing practices, and the effects of test use.
- While the evaluation of tests and personnel selection procedures depends heavily on professional judgment, the Standards provide a frame of reference to ensure that relevant issues are addressed.

Principles for the Validation and Use of Personnel Selection Procedures [Society for Industrial and Organizational Psychology (SIOP), 2003]

- These specify established scientific findings and generally accepted professional practice in the field of personnel selection psychology concerning the choice, development, evaluation, and use of personnel selection procedures designed to measure constructs related to work behavior.
- The *Principles* focus on the accuracy of the inferences that underlie employment decisions.

Application to Testing in STC Core Courses

Any test that is used as the basis for making an employee decision must:

- be based upon job information (job analysis) obtained from experts who know about the job [subject matter experts (SMEs)] who represent the population of job incumbents.
- assess components that are necessary for successful job performance such as:
 - important job behaviors.
 - knowledge, skills, and abilities (KSAs) needed to do the job.
 - work products that are representative of the job.

STC and core instructors are partners in meeting the legal and professional standards for selection procedures. STC developed the core courses in accordance with legal and professional standards by utilizing job analysis results, supplemental job research, and consultation with SMEs who were selected to represent the state. Core instructors are responsible for ensuring that tests assess the performance objectives and adhere to legal and professional selection standards. To assist in evaluating compliance with this requirement, consider the Reasonable Person Test described below. Remember, defendants (i.e., STC, agencies, core coordinators, core instructors) are responsible for providing evidence for the job-relatedness of tests.

Reasonable Person Test

Reasonable Person: An individual with considerable experience on the job who would be asked to use their professional judgment to evaluate the validity of a test.

Would a reasonable person judge the test (BSTs, WSTs, and MCTs) to be clear and unambiguous, tied to the performance objective, tied to the lesson plan, related to the job, and representative of work products?

If the answer is NO, do not use the test!

Best Practices for Testing

STC Requirements and Recommendations

The core course is a part of the selection process and agencies may base employment decisions on an individual's performance in a core course. Therefore, it is critical that the core tests represent individual performance. *For this reason, STC requires that all tests (BSTs, WSTs and MCTs) are completed individually by each trainee.*

In the event that a trainee fails a test, according to the *Standards* (AERA, APA, NCME, 1999) and the *Principles* (SIOP, 2003) opportunities for reassessment and reconsideration should be provided. It is STC's recommendation that training agencies have a remediation plan in place for reassessment. The following criteria should be incorporated into a remediation plan:

- **Multiple versions of the BSTs, WSTs, and MCTs.** When a trainee takes a test and fails and is then permitted to retake the same test as a remediation process, practice effects may influence the scores on the second test (Kaplan & Saccuzzo, 2005). When a test is given a second time, test takers score better simply because they have sharpened their skills by having taken the test the first time. For this reason, STC:
 - *recommends* that the training agency have available multiple versions of the BSTs, WSTs, and MCTs.
 - *requires* that quizzes are not the same as the final BSTs, WSTs and MCTs.
- **Communication and Coordination with the Employing Agency.** When an employing agency is informed that a trainee has not successfully completed core training, a difficult decision must be made as to whether to retain the employee. For this reason, training agencies and employing agencies should communicate and coordinate with each other especially when a trainee is struggling in a core course. This means the employing agency is aware of the remediation plan. Also, the employing agency should be consulted before a trainee is dismissed from a core course.

Edit and Review

Tests are designed to reflect course content and objectives. They are an essential component of core courses. They reinforce performance objectives and are the final measure of trainee success. Therefore, it is important to develop tests that meet these requirements. An important, often overlooked, step in developing and designing tests is a thorough editing and review process. An editing and review process can greatly improve the quality of tests (Wallach, et. al., 2006). It is STC's recommendation that core tests (BSTs, WSTs, and MCTs) are reviewed by a committee of peers prior to their first use whenever possible.

The focus of edit and review committees is to:

- detect any lack of clarity in the test (e.g., question, supporting materials, appropriate response).
- identify multiple correct answers or lack of one best answer.
- verify the correct answer.
- offer suggestions to improve the item.

For developing MCT items, it is not uncommon that for every three items that are developed, only one will make it through the review process. Consider also that a remediation test should also be developed. This means that for each MCT item that is required, at least 2 to 3 MCT item should be written. Thus, after review, the instructor will have one item for the test and another for the remediation test. A similar argument could be made for WSTs and BSTs as well.

“If item writing is considered an art, then item editing is a craft, which can improve the overall validity of a test (Baranowski, 2004, p. 349).”

Reviewing Test Performance

After tests are taken by trainees, STC recommends that instructors evaluate the performance of the tests or test items they developed. This evaluation can be as simple as determining the percent of trainees who responded appropriately to the test (BST or WST) or answered the MCT item correctly. If overall, trainees have poor performance, the instructor should determine the cause of the poor performance. Once the cause is identified, steps should be taken to correct the problem. The following are some causes of poor performance:

- the content was not included in the course instruction
- the content covered in course instruction was unclear or ineffective
- the test or item was unclear, ambiguous, or tricky
- the test or item is completely flawed

Test Security

Because core tests are part of the selection process, it is crucial that the security of the tests is ensured. A training agency's security procedures should be documented in a manual. The security procedures should address the physical security protocols for developing the test, the test answer key, printing the tests, examination sites, and inventory of testing materials (Buchendahl & Plake, 2006). In addition to adhering to a training agencies' security procedures (if one is available), instructors are responsible for ensuring the security of the tests they develop. To ensure the security of tests, the following guidelines should be followed:

- Instructors are personally responsible for and accountable for all testing material in their possession.
- Instructors must not permit trainees or other unauthorized individuals to have access to testing materials other than during test administration.
- Tests and test answers should not be left in unsecured locations (e.g., on top a desk, in a notebook, or with materials for instruction) that can easily be accessed by unauthorized individuals.
- Do not discuss tests with anyone who is not directly involved in the testing process.
- Emailing test and testing materials is not recommended. With email it is impossible to control the servers or networks on which the test may inadvertently be stored.

Structure of Multiple-Choice Items

Multiple-choice (MC) items consist of the following components:

- **Stem** – This states the question or problem. The stem may be a simple question, a statement, or a question incorporating a reading passage, chart, graph, or other information.
- **Alternatives** – These are the response options from which the test taker may choose from to answer the item.
 - The **key**, or the keyed response, is the correct alternative.
 - **Distracters** are the incorrect alternatives.



Individual Practice Exercise

For the MC items below, draw a box around the stem, circle the alternatives, and place an asterisk by the keyed response. Items were obtained from STC's item bank and Haladyna (2004).

Use the information in the table below to answer question 1.

Offender K's Schedule

Time	Activity	Duration
8:00 a.m.	breakfast	1 hour and 15 minutes
9:30 a.m.	laundry duty	1 hour
10:30 a.m.	recreation time	30 minutes
11:15 a.m.	counseling	1 hour
12:30 p.m.	kitchen duty	4 hours

1. The table above provides Offender K's schedule for the day. Which two activities will have LESS THAN a 15-minute break between them?
 - A. breakfast and laundry duty
 - B. laundry duty and recreation time
 - C. recreation time and counseling
 - D. counseling and kitchen duty

2. According to the American Film Institute, which is the greatest American film?
 - A. *It Happened One Night*
 - B. *Citizen Kane*
 - C. *Gone with the Wind*
 - D. *Star Wars*

Group Exercise I



The following two sections provide examples of common MC formats and cover the guidelines for writing multiple-choice items. These sections cover the following topics:

- Multiple-Choice Formats (pages 12 - 16)
- Multiple-Choice Item Writing Guidelines:
 - Content (pages 17 - 20)
 - Format and Style (pages 21 - 23)
 - Stem (pages 24 - 25)
 - Alternatives (pages 26 - 32)

The instructor will assign each group at least one topic (MC formats, content guidelines, format and style guidelines, stem guidelines, or alternatives guidelines). Each group will have 25 minutes to develop a five-minute presentation on their assigned topic(s) using the materials provided in the following two sections. After 25 minutes, presentations will be given to the entire group.

Multiple-Choice Formats

This section presents a variety of MC formats that are recommended by Halydyna (2004). Examples are included.

Conventional MC

The conventional MC format is the most common. There are three variations of the conventional MC format: question, incomplete stem, and best answer. An example of each is provided below.

Examples of the Three Variations of the Conventional MC Format

Question Format:

Who is John Galt?

- A. A rock star
- B. A movie actor
- C. A character in a book

Incomplete Stem Format

John Galt is a character in an Ayn Rand novel who is remembered for his

- A. integrity.
- B. romantic tendencies.
- C. courage.

Best Answer Format

Which is the most effective safety feature in your car?

- A. seat belt
- B. front air bag
- C. anti-lock braking system

Matching

The matching format is used when a set of options is available that is useful for two or more items. The format begins with a set of options at the top followed by a set of stems at the bottom. The instructions at the top inform the test takers how to respond and where to mark their answers. An example is provided below.

Example of the Matching Format

For each question select the correct answer from the options provided below.
Mark your answer on the answer sheet.

- A. Minnesota
- B. Illinois
- C. Wisconsin
- D. Nebraska
- E. Iowa

1. What is the home state of the Hawkeyes?
2. Which state is known for its cheese heads?
3. Which is the land of many lakes?
4. What state is Cornhuskers country?
5. Which state contains Cook County?

Extended Matching

The extended matching (EM) format uses a long list of options linked to a long list of stems. A set of EM items has four components: (a) a theme, (b) a set of options, (c), a lead-in statement, and (d) a set of stems. The theme focused a test taker on the context. The list of options, which can be lengthy, is possible right answers. The list of options can exhaust the content domain but must be similar in content. The lead-in statement may be a scenario or vignette. The stems should be able to be independently answered. The set of items must have at least two stems. An example is provided below. The theme for the example is neuropsychological tests.

Example of the Extended Matching Format

For each question, select the correct answer from the options provided below. Mark your answer on the answer sheet.

- A. Cognitive Estimates Test
- B. Digit Span
- C. Go-No Go Test
- D. Mini Mental State Examination
- E. National Adult Reading Test
- F. Raven's Progressive Matrices
- G. Rivermead Behavioural Memory Test
- H. Stroop Test
- I. Wechsler Memory Scale
- J. Wisconsin Card Sorting Test

A 54-year old man has a year's history of steadily progressive personality changes. He has become increasingly apathetic and appears depressed. His main complaint is increasing frontal headaches. On examination, he has word-finding difficulties. His EEG shows frontal slowing that is greatest on the left.

1. You are concerned that he may have an intracranial space-occupying lesion. Which test should you consider?
2. Tests indicate that his current performance IQ is in the low to average range. Which test should you consider?
3. The estimate of his premorbid IQ is 15 points higher than his current performance IQ. It is recommended that he has a full WAIS IQ assessment to measure both performance and verbal IQ. On the WAIS his verbal IQ is found to be impaired over and above his performance IQ. Which test is part of the WAIS verbal subtests?
4. An MRI shows a large meningioma compressing dorso-lateral prefrontal cortex on the left. Which test is most likely to indicate impairment?

Alternate Choice

Alternate choice is a conventional MC with only two options. An example is provided below. Although the alternate choice format is a downsized version of the conventional MC format, it is not a true/false format. Alternate choice requires a comparison between two choices. The true/false format does not explicitly require a comparison.

Example of the Alternate Choice Format

What is the most effective way to motivate a student?

- A. intermittent praise
- B. consistent praise

True/False

The true/false (T/F) format has been well established in classroom assessment but is seldom used in standardized testing programs (e.g., SAT, GRE). T/F format can be traditional or unusual. An example of each is provided below. The T/F format should be used with caution because research suggests it tends to be used to assess recall of trivial knowledge.

Examples of the T/F Format

Traditional T/F Format:

Mark A on your answer sheet if true and B if false.

1. The first thing to do with an automatic transmission that does not work is to check the transmission fluid.
2. The major cause of tire wear is poor wheel balance.
3. The usual cause of clutch “chatter” is in the clutch pedal linkage.
4. The distributor rotates at one half the speed of the engine crankshaft.

Unusual T/F Format:

Place an X beneath each structure for which each statement is true.

Characteristic	Root	Stem	Leaf
Growing point protected by a cap			
May possess a pithy center			
Epidermal cells hair-like			
Growing region at tip			

Multiple True/False

The multiple true/false (MTF) format has much in common with the TF format. The difference is that TF items should be different in content. However, MTF has similar content that is a result of a lead-in statement. For this reason, it is also similar to the EM format. The MTF format can be simple or complex. An example of each is provided below.

Examples of the MTF Format

Simple MTF Format:

Mark A on your answer sheet if true and B if false.

The Lion, the Witch, and the Wardrobe by C. S. Lewis can best be summarized by saying:

1. A penny saved is a penny earned.
2. If you give them an inch, they will take a mile.
3. Good will always overcome evil.
4. Do not put off tomorrow what you can do today.
5. Do not put all your eggs in one basket.

Complex MTF Format:

Your video store rents VHS for \$2.00 on weekdays and \$3.00 on weekends. You also rent DVDs for \$3.00 on weekdays and \$4.00 for weekends. Below is a weekly summary of rentals.

	Videos Rented	
	VHS	DVDs
Monday	38	35
Tuesday	31	28
Wednesday	40	45
Thursday	47	49
Friday	55	52
Saturday	63	60
Sunday	75	68

Mark A on your answer sheet if true and B if false.

1. The video store makes more money from VHS than from DVD.
2. DVDs and VHSs are more expensive on the weekdays.
3. The video store rents more DVDs in a week than VHS.
4. DVDs are more expensive than VHS.
5. The video store rents more videos Friday, Saturday, Sunday than on the weekdays.
6. Customers rent about the same number of DVDs and VHSs on the weekdays.
7. The video store rents more VHS than DVDs on the weekends.

Multiple-Choice Item Writing Guidelines

Many books and journal articles have been written that provide guidance on writing MC items. However, despite this guidance, item writing is not yet a science. It is still part art and science. The guidelines presented in this section were quoted from and adapted from a journal article by Haladyna, Downing, and Rodriguez (2002) and *Developing and Validating Multiple-Choice Test Items* (Haladyna, 2004, p. 97 - 121). These guidelines were drawn from textbooks and analyses of over 90 research studies on the validity of these guidelines.

Content Guidelines

I. Assess the Core Course Performance Objective

Items must be carefully written so that they provide a valid indicator of the trainee's ability to meet the associated performance objective. The verb(s) in the performance objective has a direct effect on the test item. The item must directly measure the performance specified by the verb in the performance objective.

Example of Assessing the Performance Objective

Performance Objective: Identify the primary legal references and sources that impact the work of a corrections officer. For example: code statutes, case law, and regulations.

Poor MC Item:

Which of the following agencies is responsible for establishing and maintaining the minimum standards for local corrections facilities?

- A. Office of Administrative Law
- B. Corrections Standards Authority
- C. State Personnel Board
- D. California Department of Corrections and Rehabilitation



Better MC Item:

Which of the following options is a primary legal reference or source that impacts the work of a corrections officer?

- A. Title 15
- B. Evidence Code
- C. Title 2
- D. Civil Code



In the example provided above, the performance objective specifies that the trainee will identify primary legal references and sources. The poor MCT item is inappropriate because the question is not related to the performance objective. The better MCT is more appropriate because it requires test takers, who are given a list of possible legal references and source, to identify the option that impacts the work of a corrections officer. The better MCT measures the performance objective.

2. Assess Important Content

MC items should only assess important content and avoid trivial content. Judging the importance of content is subjective. Haladyna (2004) suggests consulting with other content experts to determine if the content is too trivial. The example below shows the difference between trivial and important content. The poor item asks for a fact that is trivial to trainees. The better item addresses the population for which a specific risks and needs assessment was developed. The better item seems more appropriate for probation officers.

Example of Assessing Important Content

Poor MC Item:

When was the STRONG created?



Better MC Item:

The STRONG was created as a risks and needs assessment for which of the following populations?



3. Use Novel Material to Test for Understanding

It is important to test for understanding rather than simple recall. Testing understanding can be done by presenting a concept, procedure, or principle in a novel way. This can be done by presenting scenarios or vignettes that require the trainee to use critical thinking skills or problem solving. The example below shows the difference between testing recall and understanding. The poor item provides definitions which are likely to be memorized. The better item tests understanding. The trainee who understands a metaphor will select the correct sample.

Example of Using Novel Material to Test for Understanding

Poor MC Item:

Which is the best definition of a metaphor?

- A. Metaphors describe something as if they are something else.
- B. Metaphors make comparisons to other things.
- C. Metaphors are trite; stereotyped expressions.



Better MC Item:

Which of the following is a metaphor?

- A. At the breakfast buffet, I ate like a pig.
- B. My cat has fur like knotted wool.
- C. She is like a rose, full of thorns and smelly.



4. Keep Content of Each Item Independent

When writing multiple items, it is possible to provide information in one item that will help the test taker correctly answer another item. Test wise trainees will look for clues in items. For example, consider the two items below. A test taker who correctly answers the first item can use the information to assist with answering the second item. Because Roxie is correct for the first item, it cannot be for the second. Also, because Kate was mentioned in the first item, it can be ruled out as the correct answer. Therefore, the correct answer must be Sara.

Example of Item Dependence

MC Item:

Who was Lupe's best friend?

- A. Kate
- B. Delores
- C. Roxie*



MC Item:

Who was quarreling with Lupe?

- A. Kate
- B. Sara*
- C. Roxie

5. Avoid Overly Specific or Overly General Content

Items may assess knowledge on a continuum from very specific to very general. Items should be written with this continuum in mind. That is, the extremes should be avoided. Overly specific items tend to be trivial and overly general items tend to be ambiguous. In the examples below the first item is too specific and the second is too general.

Example of Very Specific and Overly General Content

Very Specific Item:

Who wrote the Icon of Seville?

- A. Lorca
- B. Ibanez
- C. Rodriquez

General Item:

Which is the most serious problem in the world?

- A. hunger
- B. lack of education
- C. disease



6. Avoid Opinion-Based Items

Items should assess well-known facts, concepts, principles, and procedures. It is unfair to test a trainee on an opinion unless that opinion is qualified by analysis, supported by evidence, or presented during instruction. The example below shows the difference between an unqualified and qualified opinion.

Example of Avoiding Opinions

Unqualified Opinion Item:

Which is the best comedy film ever made?

- A. Abbot and Costello Go to Mars
- B. Young Frankenstein
- C. A Day at the Races

Qualified Opinion Item:

According to the American Film Institute, which is the greatest American film?

- A. It Happened One Night
- B. Citizen Kane
- C. Gone with the Wind
- D. Star Wars

7. Avoid Trick Items

Trick items deceive the test taker into choosing a distracter instead of the right answer. Items may be written to deliberately trick the test taker or may accidentally trick the test taker. Test takers perceive the following elements as tricky:

- The item appears to intentionally deceive, confuse, or mislead.
- Trivial content was presented.
- The difference between the alternatives was too fine.
- The item included information that was not important to the problem.
- Principles were presented differently than in instruction.
- The item was too ambiguous such that even the best trainee didn't know the answer.

According to Haladyna (2004) trick items are difficult to illustrate, but the open-ended questions below are a few examples of trick items. Of course, there is a fourth of July in England, as there is around the world. Also, all months have 28 days.

Examples of Trick Items

Is there a fourth of July in England?

Some months have 31 days. How many have 28?

Format and Style Guidelines

8. Use Question, Phrase, or Best Answer Format

The most common MC item formats are question, incomplete stem, and best answer. Each format has the following three parts: (a) a stem, (b) the correct choice, and (c) several wrong answers called distracters. An example of each is provided below. The use of these MC item formats is recommended by Haladyna, Downing, and Rodriguez (2002) based on a review of various item formats that were identified as supported, not supported, or uncited in the text-books that were reviewed.

Examples of Common MC Formats

Question Format:

Who is John Galt?

- A. A rock star
- B. A movie actor
- C. A character in a book

Incomplete Stem Format

John Galt is a character in an Ayn Rand novel who is remembered for his

- A. integrity.
- B. romantic tendencies.
- C. courage.

Best Answer Format

Which is the most effective safety feature in your car?

- A. seat belt
- B. front air bag
- C. anti-lock braking system

9. Format Items Vertically

Horizontal formatting is sometimes preferred because more items can fit on a single page. However, vertical formatting is more visually appealing because it looks less cramped. Students with test anxiety may find horizontal formatting harder to read. As a result, it may confuse students and lower test scores. The examples below illustrate horizontal and vertical formatting.

Example of Horizontal and Vertical Format

Horizontal Item Format:

You draw a card from a deck of 52 cards. What is the chance you will draw a card with an odd number on it?

- A. $36/52$ B. $32/52$ C. About one half



Vertical Item Format:

You draw a card from a deck of 52 cards. What is the chance you will draw a card with an odd number on it?

- A. $36/52$
B. $32/52$
C. About one half



10. Edit and Proof Items

Clarity is the most important consideration in editing and proofing items. The first step for editing and proofing items is determining if the central idea is presented clearly. Once the item is edited for clarity, editing to ensure that each item has correct grammar, punctuation, capitalization, and spelling is also important.

11. Use Appropriate Vocabulary and Reading Level

MC items are developed to measure knowledge and skills. Some items can even measure aspects of problem solving, critical thinking, and other cognitive abilities. Unless the test is measuring reading comprehension, the test taker's reading level should not affect test performance. Thus, vocabulary and reading level should be simple and at the level of the weakest readers in the group. Unless reading comprehension is a focus of the tests, items should not measure both reading comprehension and the desired knowledge or skill.

12. Minimize Reading Time

Items should not be overly wordy. In fact, wordy items tend to lack clarity. Long test items will also require longer administration time. If a fixed testing time is available, wordy items limit the number of items that can be presented. With fewer items, the content area cannot be sampled appropriately. For these reasons, items should be as brief as possible without compromising the item. Both the stems and the alternatives should get to the point and be as brief as possible. In the example below, the first item has repetitive wording. The second item eliminates the unnecessary words.

Example of Minimizing Reading Time

Wordy Item:

Effective student grading should probably minimize:

- A. the student's present status.
- B. the student's progress against criteria state in the syllabus.
- C. the student's status relative to his or her ability.
- D. the student's status relative to the class.
- E. the student's progress relative to state standards.



Better MC Item:

Which should determine a student's grade?

- A. Achievement against stated criteria.
- B. Status relative to other class members.
- C. Progress relative to his or her ability.



13. Proofread and Have Each Item Peer Reviewed

The most important part of item writing is proofreading. Preferably the items are proofread by a peer. Thus, they can be reviewed not only for correct grammar, formatting, spelling, and capitalization, but also for clarity. Even after thorough review, tests will have items with errors. According to Haladyna (2004), if four errors are found in the final review of an item, at least one error was not found. Errors in a test suggest carelessness and negligence. This may lead to the impression of poor test development. This should not be conveyed to a test taker. Additionally, errors may reduce clarity in the item and may distract test takers resulting in lower scores.

Stem Guidelines

14. Clear Directions

The stem should be written so that the test taker knows the focus of the item immediately. From the stem the test taker should know what situation or problem is presented. In the poor stem below, a test taker would have to review the options to guess what happened to the plant. In the better stem, the clearer directions lead to the expectation that the plant would grow in the direction of the light source.

Example of Clear Directions

Poor Stem:

A plant in a flower pot fell over. What happened?



Better Stem:

A plant growing in a flower pot was turned on its side.
A week later, what would you expect to see?



15. Include Central Idea

The item stem should always include the main idea. The test taker should always know what is being asked after reading the stem. Failing to put the main idea in the stem may affect the quality of the item. In the example below the poor item does not provide adequate information while the second is more direct, asks a question, and provides plausible options.

Example of Including the Central Idea

Poor Item:

Corporal Punishment:

- A. has been outlawed in many states.
- B. is psychologically unsound for school discipline.
- C. has many benefits to recommend its use.



Better Item:

What is corporal punishment?

- A. a psychologically unsound form of school discipline
- B. a useful disciplinary technique if used sparingly
- C. an illegal practice in our nation's schools



16. Make as Brief as Possible

Lengthy items increase administration time. Thus, it is a good idea to keep the stems as brief as possible while still including all of the necessary information. The example below shows both lengthy and brief stems.

Example of Brevity

Lengthy Stem:

Destruction of certain cortical tissue will lead to symptoms affecting our behavior. However, this destruction may lead to symptoms that are due to withdrawal of facilitation in other cortical areas. Thus, the tissue damage affects cortical functioning both directly and indirectly. While such effects may be temporary, what is the typical recovery time?

- A. immediately
- B. several days to a week
- C. several months
- D. seldom ever



Brief Stem:

What is the typical recovery time for cortical functioning when tissue is destroyed?

- A. immediately
- B. several days to a week
- C. several months
- D. seldom ever



17. Word Positively

The use of negative words in the stem (e.g., not, except, etc.) has negative effects on test takers and their performance. Test takers may have difficulty understanding the meaning of negatively phrased stems. The example below shows the use of **EXCEPT** in a stem. While this is a popular item format, it is not recommended. This type of item taxes a test taker's working short-term memory. If a negative term is used, it should be in bold type, underlined, capitalized, or all three of these. This is because test takers generally read through the **NOT** and forget to reverse the logic required to answer the question. This is why **NOT** shouldn't be used in stems.

Example of Using "Except"

Each of the following drugs is appropriate for the treatment of cardiac arrhythmia EXCEPT one. Which one is the exception?

- A. Phenytoin
- B. Lidocaine
- C. Quinidine
- D. Propranolol
- E. Epinephrine



Alternative Guidelines

18. Develop as Many as Possible, but Three is Adequate

Research is beginning to support the use of three alternatives for MC items. That is, two distracters and one correct answer. However, there is a slight benefit to having as many alternatives as possible, but only if they are plausible and different from each other. As many alternatives as possible should be written, but don't be surprised when only one or two work as intended.

19. Only One Correct Choice

Unless the whole test is designed to have test takers choose the best of a correct set of alternatives, and this is virtually never done, then make sure that there is only one choice that is technically correct.

20. Balance the Key Position

Item writers tend to place the key as the B or C response options. Also, wise test takers will identify patterns for key responses and use this information when they are in doubt as to the correct answer. For these reasons, the location of the key should be varied. For example, consider a 100-item test and each item has four alternatives. Approximately 25% of the items should have A as the correct option or keyed response. Thus, 25% of the items should have B as the correct option. This applies to options C and D as well.

21. Logical Order

The alternatives should always appear in logical or numerical order. The example below shows two versions of the same item, first with the numerical order wrong and then with the order correct. Alternative should always be in either ascending or descending order. Also, decimal points should always be aligned. The intent of an item is to test for knowledge, not have the test taker hunt for a correct answer. Doing so only wastes the test taker's time and causes unnecessary stress.

Example of Numerical Order

Incorrect Numerical Order:

What is the cost of an item that normally sells for \$9.99 that is discounted 25%?

- A. \$5.00
- B. \$7.50*
- C. \$2.50
- D. \$6.66



Correct Numerical Order:

What is the cost of an item that normally sells for \$9.99 that is discounted 25%?

- A. \$2.50
- B. \$5.00
- C. \$6.66
- D. \$7.50*



Logical order is more difficult to demonstrate. Logical order of items requires judgment, but alternatives may be ordered based on the order of the information provided in the stem, alphabetically, or by the form of the alternatives.

Example of Logical Order

Alphabetical Order:

Which is the most important consideration in preparing a waxajet?

- A. lubricant*
- B. O-ring integrity
- C. positioning
- D. wiring



Organized by Length:

When an item fails to perform on a test, what is the most common cause?

- A. The item is faulty. *
- B. Instruction was ineffective.
- C. Student effort was inadequate.
- D. The item failed to match the objective.



22. Keep Choices Independent

Alternatives should not be overlapping. These types of options may give clues to wise test takers about the correct and incorrect response. Also, if alternatives have overlapping values or meanings, it is possible that the item has more than one correct answer. Consider the example below. If the correct answer is 25 years of age, then both C and D are correct. This can be corrected by creating ranges that are distinct from each other. These types of errors can lead to embarrassing challenges to test items.

Example of Non-Independent Choices

What age range represents the physical “peak” of life?

- A. 11 to 15 years of age
- B. 13 to 19 years of age
- C. 18 to 25 years of age
- D. 24 to 32 years of age
- E. over 32 years of age



23. Similar in Content and Grammatical Structure

Alternatives should be similar in content. Alternatives that are different in content provide a clue to the test taker. These cues are not the intent of the item and are an unfortunate accident. To avoid this situation, the alternatives should be similar in content. In the example provided below, the correct answer, B, is different from the other alternatives. It is the only alternative that mentions beta burns and therefore provides a clue to the test taker.

Example of Content Clues

The treatment of Beta burns depends on the nature and seriousness of the injury. If no physician is available, Beta burns should be treated just like other burns. After all fallout is removed, such burns should be bandaged, and salves and jellies should be used to protect the wound.

According to the paragraph, which statement is most correct?

- A. All minor burns should be treated by a physician.
- B. Beta burns should be treated similar to heat burns.
- C. Radioactive burns should be covered with salves and jellies for protection.
- D. Fallout should be removed from all burns.



Alternatives should be similar in grammatical structure. Alternatives that have different grammatical structure provide a clue to the test taker. Several examples are provided below. In the first example because the stem ends with “an,” the correct response must begin with a vowel. Thus, even a test taker who does not know the answer could use the clues to answer the item correctly. In the second example, the article “a” in the stem indicates that the correct response is singular. Therefore, a wise test taker could rule out A, C and D as correct answers.

Example I of Grammatical Clues

Poor Item:

An example of an extinct animal is an:

- A. sloth
- B. Eskimo curlew
- C. gorilla
- D. baboon



Better Item:

Which of the following animals is extinct?

- A. sloth
- B. Eskimo curlew
- C. gorilla
- D. baboon



Example 2 of Grammatical Clues

Poor Item:

Penicillin is obtained from a:

- A. mold
- B. bacterium
- C. coal-tars
- D. tropical trees



Better Item:

From which of the following is penicillin obtained?

- A. mold
- B. bacterium
- C. coal-tars
- D. tropical trees



24. Equal Length

The length of the alternatives should be kept about the same. Item writers tend to make the correct choice the longest. This should be avoided. Generally, this occurs because the stem and correct answer are written and then the distracters are written quickly to complete the item. In the example below the key response for the poor item, D, is longer and more detailed than the other alternatives. A test taker without knowledge of the subject matter would likely choose D because it is more detailed than the other options. The better item fixes this problem.

Example of Length of Alternatives

Poor Item:

Which of the following organizational development interventions is designed to improve the effectiveness of individuals?

- A. process consultation
- B. team building
- C. survey feedback
- D. education and training designed to increase skills in the areas of decision-making, problem-solving, and goal setting



Better Item:

Which of the following organizational development interventions is designed to improve the effectiveness of individuals?

- A. process consultation
- B. team building
- C. survey feedback
- D. skills training



25. Avoid “None of the Above” and Never use “All of the Above”

It is easy to use “none of the above” as a final alternative. However, the research on the usefulness of these options is mixed and controversial. Some researchers support the “none of the above” format in that it forces test takers to solve the problem rather than choose the correct answer. Thus, it may be useful for math problems. “None of the above” should be used sparingly. If it is used, it should be the correct answer an appropriate number of times. For example, if 4 items use “None of the above as an alternative. At least one of the items should have “None of the above” as the correct response.

Never use “all of the above.” The use of this as an alternative is controversial. Item writers sometimes use it because it is easy to identify one, two, or even three correct answers. But using this choice may help wise test takers. That is, a test taker with partial knowledge may identify two choices as correct, but not the third. The test taker could then choose “all of the above” as the correct choice. The use of “all of above” seems to assess test taking strategies rather than the knowledge the item was designed to measure. For this reason, it should be avoided.

26. Phrase Positively

Just as stems should be phrased positively, so should the alternatives. The use of negatives such as NOT and EXCEPT should be avoided. These words have negative effects on test takers and their performance. Test takers may have difficulty understanding the meaning of negatively phrased sentences.

27. Avoid Giving Clues to the Correct Answer

The following clues tip off test takers to the correct answers:

- **Specific determiners:** Specific determiners include the terms always, never, totally, absolutely, and completely. These are so extreme that they are seldom the right answer. If the use of one of these is appropriate for the right answer, be sure at least one of the correct alternatives also contains one.
- **Clang associations:** Clang associations occur when a word or phrase is used in the stem and it also appears in an alternative. Usually this alternative is the right answer. If it is not, the item might be a trick question.
- **Blatantly absurd, ridiculous options:** When trying to develop a third or even fourth alternative it is tempting to develop an absurd or ridiculous choice as either humor or out of frustration. In either case these alternative will not be chosen. Distracters that are not plausible are useless.

An example of each clue is provided below.

Examples of Giving Clues

Item with a Specific Determiner:

Which of the following does research on homework support?

- A. Never assign homework on Fridays.
- B. Homework should be consistent with class learning.
- C. Always evaluate homework the next day.



Item with Clang Association (tax):

What is the purpose of the tax table? To help you determine:

- A. your gross income.
- B. the amount of tax you owe.
- C. your net earnings.
- D. your allowable deductions.



Item with Ridiculous Distracters:

Who is best known for contributions to microelectronics?

- A. Comedian Jay Leno
- B. Robert Sveum
- C. Actor Bruce Willis



28. Make all Distracters Plausible

Items are used to measure knowledge and skills. Thus, the key must be right and the distracters wrong. But, the wrong answers must be plausible. Plausibility refers to the idea that the test takers with a high degree of knowledge should answer the item correctly. However, test takers with little knowledge should answer the item incorrectly. A plausible distracter will look right to a test taker who does not possess the required degree of knowledge. Thus, the incorrect options “distract” the less knowledgeable test taker.

Once a quality stem is written, the effectiveness of a test item depends on the quality of the distracters. The series of examples provided on the following page shows this principle. In the item with very poor distracters, a test taker with no knowledge of history could correctly answer the item. This is because the only plausible option is C, Franklin D. Roosevelt, because the remaining options are a well-known cartoon character, an actor, and an author. The item with poor distracters is a slight improvement because at least all of the options are or were presidents. However, even test takers with limited knowledge could rule out B, Barack Obama, and C, George Washington. Obama can be ruled out because he is the current president. George Washington can be ruled out because most people know he was the first president. A test taker with limited knowledge could rule out Andrew Jackson because he was the 7th president and known for creating the modern Democratic Party; which must have been a long, long time ago. The better item has the best distracter as the options are all presidents who were not very recent or very early presidents. In fact, the options were the 30th, 31st, 32nd, and 33rd presidents.

Examples of Distracter Plausibility

Item with Very Poor Distracters:

Who was the thirty-second United States President?

- A. Mickey Mouse
- B. David Hasslehoff
- C. Franklin D. Roosevelt
- D. Anne Frank



Item with Poor Distracters:

Who was the thirty-second United States President?

- A. George Washington
- B. Barack Obama
- C. Franklin D. Roosevelt
- D. Andrew Jackson



Item with Better Distracters:

Who was the thirty-second United State President?

- A. Calvin Coolidge
- B. Herbert Hoover
- C. Franklin D. Roosevelt
- D. Harry S. Truman



29. Use Typical Errors of Students to Create Distracters

Good plausible distracters come from a thorough understanding of common trainee errors. These can be identified when a trainer reviews student work products (e.g., reports, documents students write, answers to essay or short answer questions, scratch paper). In these work products, trainees may provide the correct answer and plausible wrong answers that are actual trainee errors. In the example below, distracter A is an incorrect answer and represents an error commonly made by someone who is just learning addition. An individual who forgets to carry the '1' would obtain A as the answer.

Example of Distracter with Typical Student Errors

$77 + 44 =$

- A. 101
- B. 111
- C. 112



Group Exercise 2



Now that you are familiar with the guidelines for writing multiple-choice items, each group will practice writing an item that adheres to the guidelines and assesses a performance objective. This group exercise consists of two segments: development and peer review. The instructions for each segment are outlined below.

Development

1. Each group has 15 minutes to develop a multiple-choice item.
2. Select one learning objective from the options provide on the following page. The performance objectives below were selected directly from the STC Core Manuals. Five objectives are provided from each core manual (probation offer, adult corrections officer, juvenile corrections officer). Each group only needs to select one objective.
3. Write a multiple-choice item that assesses the chosen performance objective and adheres to the item writing guidelines.
4. Using the envelope provided:
 - a. write the objective and team name at the top of the envelope.
 - b. write the multiple-choice item below the objective.

Peer Review

1. Each team envelope will be passed to the other teams in a counter-clockwise direction.
2. Each team will have 5 minutes to answer and review the multiple-choice item on the envelope they received. Using the index cards provided as a team record:
 - a. the answer to the multiple-choice item (e.g., A, B, C, or D).
 - b. written feedback that focuses on the item writing guidelines.
3. Repeat steps 1 thru 3 until each group receives their item back (approximately 15 minutes).
4. Each group will take 5 minutes to review the feedback received for the item.

ACO Performance Objectives

- Distinguish the difference between a felony, a misdemeanor, an infraction and a wobbler.
- Identify the reasons why appropriate classification of offenders is important.
- Identify the tests that evidence must successfully pass before it may be entered into criminal court (Evidence Code Section 210) including must:
 - be relevant to the matter at hand.
 - be competently presented in court.
 - have been legally obtained.
- Describe principles of effective communication in a jail setting.
- Identify basic information that can be disclosed about offenders.

JCO Performance Objectives

- Generate a description of the common behavioral characteristics, patterns, and physical symptoms associated with common drugs used by juveniles.
- Identify the major personality, behavioral, and physical indicators associated with child abuse.
- Describe juvenile's actions that may lead to assaultive behavior.
- Explain the limitations for use of handcuffs.
- Identify the requirements for monitoring a juvenile who is placed in a safety room, including the following:
 - direct visual supervision.
 - time intervals for documenting juvenile's behavior and any staff interventions.
 - medical evaluation.
 - mental health evaluation.

PO Performance Objectives

- Student will be able to identify the appropriate action to take when made aware of unethical and/or criminal conduct of other probation officers, correctional professionals, or others.
- Given various scenarios, the student will be able to identify situations where a duty to inform exists.
- Student will be able to identify guidelines and principles for interacting with suspected gang members.
- Student will be able to identify special considerations in assessing and managing sex offender cases.
- Student will be able to identify situations when the use of restraints may not be appropriate.

Group Exercise 3

Group 1 reads page 36 and reports out on “Solicit Appropriate Trainee Responses”

Group 2 reads page 37 and reports out on “Measure the Core Course Performance Objective”

Group 3 reads page 38 and reports out on “Provide the Resources specified in the Performance Objective”

Group 4 reads page 39 and reports out on “Assess Important Content”

Guidelines for Designing BSTs and WSTs

Solicit Appropriate Trainee Response

The response that the test solicits from the trainee should be appropriate for the test type.

- BST should require the trainee to perform a physical behavior that can be seen (e.g., interview, handcuff, demonstrate a takedown, etc.). BSTs should not result in a test that has a trainee write a report, justification for a response, etc.
- WSTs should require the trainee to produce a written document such as a report, completed form, case plan, etc.
- BSTs and WSTs are to be completed individually by each trainee.

BST Example

Performance Objective: Demonstrate how to safely approach a compliant subject in a standing, kneeling, and prone position prior to the application of handcuffs. (BST)

Poor BST: The trainee is provided with several diagrams that illustrate several options for approaching a subject in a standing position. The trainee is asked to circle the diagram that corresponds to the safest approach for handcuffing the subject.

Better BST: The trainee is presented with a person in the standing position. The trainee must physically demonstrate in front of the instructor the appropriate safe approach for the application of handcuffs. This procedure is then repeated with the trainee approaching a person in both kneeling and prone positions.

In the example provided above, the performance objective specifies that the trainee must demonstrate how to safely approach a compliant subject prior to the application of handcuffs. The poor BST is inappropriate because it does not require a physical demonstration of the required behavior; that is a safely approaching a compliant subject prior to the application of handcuffs. It also does not require the trainee to demonstrate the behavior for the three positions (standing, kneeling, and prone). For these reasons, the better BST is more appropriate as it requires a physical, visible behavior.

Measure the Core Course Performance Objective

Each BST and WST must be carefully designed so that it provides a valid indicator of the trainee's ability to meet the associated performance objective. The verb(s) in a performance objective has a direct effect on the test design. The test must directly measure the performance specified by the verb(s) in the performance objective.

WST Example

Performance Objective: Given a sample intake profile and a sample classification plan, classify the juvenile and assign housing according to key classification factors, and justify your decision. (WST)

Poor WST: A written skills test that asks the trainee to list key factors that should be considered in classifying and assigning a juvenile to a housing unit.

Better WST: A written skills test that provides a sample intake profile and classification plan. The test asks multiple questions requiring the trainee to write a narrative that (a) classifies the individual, (b) assigns housing, and (c) justifies the classification and housing assignment of the juvenile.

In the example provided above the key verbs in the objective are classify, assign, and justify. The poor WST is inappropriate because it does not require the trainee to apply knowledge of classification and housing assignments. It also does not require the trainee to justify the classification or housing assignment. For these reasons, the better WST is more appropriate as it is directly associated with all of the verbs in the performance objective.

Provide the Resources Specified in the Performance Objective

BSTs and WSTs require resources specified in the condition of the performance objective. These resources may include, but are not limited to, sample scenarios, penal code, handcuffs, sample case plans, sample procedures, etc. Instructors may need to develop these resources as part of designing the BST or WST. For WSTs the vocabulary and reading level of the resources should be at the appropriate level for the trainees.

WST Example

Performance Objective: Presented with a scenario of offender behavior, generate a written report of the offender's behavior and appearance. Include the following:

- factors that indicate risk
- statements made by offender
- corrections officer actions

Poor WST: A written skills test that asks the trainee to list content that should be included when documenting an offender's behavior or appearance.

Better WST: A written skills test that provides a scenario (e.g., written description, vignette, video, or skit/play) that includes an offender whose behavior or appearance requires documentation. The test question requires the trainee to write a narrative that (a) documents the offender's behavior (if necessary), (b) appearance (if necessary), and (c) the required content as identified in the performance objective.

In the example provided above, the performance objective specifies that the trainee will be provided with a scenario of offender behavior. The poor WST is inappropriate because it does not provide a scenario for the trainee to use in the assessment of the performance objective. It also does not require the trainee to apply job knowledge. The trainee is simply asked to list what should be included in documentation, rather than actually writing a report that includes the required content based on a provided scenario. For these reasons, the better WST is more appropriate because it provides the resources specified in the performance objective and requires the application of job knowledge.

Assess Important Content

BSTs and WSTs should measure important content from the course that is necessary to perform the job. The tests should not measure trivial knowledge, skills, or abilities that are not pertinent to successful job performance. It is important to remember that these tests are part of the selection process. Trainees should not fail based on irrelevant, trivial content that is not job-related or critical to job performance.

WST Example

Performance Objective: List key considerations that must be addressed in planning cell extractions.

Poor WST: A written skills test that asks the trainee to describe a cell extraction incident that did not go as planned.

Better WST: A written skills test that requires the trainee to write a list of the key considerations that must be included in a cell extraction plan, which includes the required content presented during the instruction of the course module.

In the example provided above the performance objective specifies that the trainee will list key considerations for planning cell extractions. The poor WST is inappropriate because it does not measure the performance objective. Telling a story about a poorly executed cell extraction may be considered trivial compared to identifying the key considerations for planning a cell extraction. The better WST is more appropriate because it measures the performance objective rather than knowledge of a past incident.

Pass the Reasonable Person Test

Would a reasonable person judge the test (BST, WST, or MCT) to be clear and unambiguous, tied to the performance objective, tied to the lesson plan, related to the job, and representative of work products? If the answer is NO, do not use test. The test should be redesigned before it is used to measure the associated performance objective.

BST Example

Performance Objective: Demonstrate proper manipulation of handcuffs. For example:

- retrieval
- pistol grip by chain
- ensuring swing arm is facing toward subject's wrist
- ensuring that handcuffs are not double locked before attempting to place them on the subject's wrists
- once placed on subject's wrists, check for excessive tightness before double locking
- double locking on subject's wrists

Poor BST: Ask the trainees to raise their hands if they already learned how to correctly apply handcuffs in the PC 832 course. Have only those students who did not raise their hand demonstrate the proper manipulation of handcuffs.

Better BST: Require each trainee to correctly apply handcuffs on a subject incorporating the principles identified in the performance objective.

In the example provided above, the performance objective specifies that the trainee will demonstrate the proper manipulation of handcuffs. The poor BST is inappropriate because it does not require each student to demonstrate the physical task of correctly applying handcuffs to a subject. Simply asking students if they learned it in another training class is inappropriate. A reasonable person could speculate that a trainee saying he or she took a previous course that included handcuffing does not indicate that the trainee can definitely apply handcuffs correctly. Therefore, the better BST is more appropriate because it requires each trainee to physically apply handcuffs correctly to a subject, incorporating the principles outlined in the performance objective.

References

- Age Discrimination in Employment Act, 29 U.S.C. §§ 621 et seq. (1967).
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington D.C.: Author
- Americans with Disabilities Act, 42 U.S.C. §§12101 et seq. (1990).
- Baranaowski, R. A. (2006). Item editing and editorial review. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp.349-357). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Buckedahl, C. W. & Plake, B. S. (2006). Evaluating Tests. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp.349-357). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Cal. Pen. Code § 6035.
- Civil Rights Act of 1964, 42 U.S. Code, Stat 253 (1964).
- Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice (1978). Uniform guidelines on employee selection procedures. *Federal Register*, 43(166), 38290-38309.
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309-334.
- Kaplan, R. M., & Saccuzzo, D. P. (2005). *Psychological Testing: Principles, Applications, and Issues* (6th ed.). Belmont, CA: Thomson Wadsworth.
- Society for Industrial and Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.). Bowling Green, Ohio: Author.
- Title 15 CCR § 131.
- Wallach, P. M., Crespo, L. M., Holtzman, K. Z., Galbraith, R. M., & Swanson, D. B. (2006). Use of committee review process to improve the quality of course examinations. *Advances in Health Sciences Education*, 11, 61-68.

Additional Resources

Uniform Guidelines on Employee Selection Guidelines –
<http://uniformguidelines.com/uniformguidelines.html>

Uniform Guidelines Questions and Answers –
<http://uniformguidelines.com/questionandanswers.html>

SPB Selection Analyst Training Program –
http://www.spb.ca.gov/training/catalog/selection_analyst.htm

WRIPC (Western Region Intergovernmental Personnel Assessment Council) occasionally has one-day item writing courses –
<http://www.wripac.org/>