

HANDBOOK OF LEARNER EVALUATION & TEST ITEM CONSTRUCTION

*Adapted from the Handbook written for the
OMERAD Primary Care Faculty Development Fellowship Program*

Office of Medical Education Research and Development
College of Human Medicine
Michigan State University

December 12, 2011



Content Development

Christopher B. Reznich, PhD
Stephen L. Yelon, PhD

Handbook Development

Christopher B. Reznich, PhD

Page Design

Deborah A. Sleight, PhD
Molly M. Tranberg

Table of Contents

Introduction	1
PART 1: Qualities of All Good Tests.....	3
Purposeful	3
Valid.....	3
Reliable.....	3
Objective.....	4
Comprehensive.....	4
Differentiating	4
Expected.....	4
Instructive	5
Useful	5
PART 2: General Standards for Test Items	7
PART 3: Technical Specifications for Types of Test Items	11
True or False Questions	11
Multiple Choice Questions	13
Matching Items Questions	16
Completion Questions	19
Short Answer Questions	20
Essay Questions.....	20
Production Questions	21
Procedures: Criterion Checklists	24
References.....	27
Appendix: Test Item Worksheets.....	29
True/False Questions	31
Multiple Choice Questions	33
Matching Questions	35
Completion Questions	37
Short Answer Questions	39
Essay Questions.....	41
Production Questions	43
Procedures: Criterion Checklists	45

Introduction

This handbook was written to help you produce effective tests and test items to use with your residents and medical students. It was originally written for the OMERAD Primary Care Faculty Development Fellowship Program.

In this manual you will learn how to create tests and test items that are honed, precise and accurate. There are three major sections in this manual:

1. Qualities of all good tests
2. General standards for test items
3. Technical specifications for types of test items

In the first section you will learn the specific qualities of tests that make them good. In the second section you will look at examples of test items and how they should be written to avoid ambiguity. The third section contains guidelines for each type of test item (true/false, multiple choice, matching, completion, short answer, essay, production and procedure). These will tell you where and why these types of test items are used and how to write and use them properly.

To make the most out of this manual you should have at your side the test items you have already written so you can check them against the guidelines. If your items don't meet the guidelines, study the qualities, general standards and technical specifications then rewrite the test items accordingly.

PART 1: Qualities of All Good Tests

What makes for a good test? Good tests have certain qualities that make them good. All good tests are purposeful, valid, reliable, objective, comprehensive, differentiating, expected, instructive and useful. You will want to keep these qualities in mind any time you create or revise a test.

Qualities of a good test are distinct from desirable qualities for individual test items, which will be addressed later.

Purposeful

A good test serves a purpose, either formative, to help learners improve their performance or knowledge, or summative, to certify that learners have indeed learned what they should have learned.

Formative evaluation usually comes during the course of instruction, such as quizzes taken at the end of a rotation that are intended to show learners where they need to improve. Summative evaluation usually comes at the end of the course of instruction. Board certification – the knowledge and performance tests doctors take for certification – is a form of summative evaluation. So any performance evaluation is intended to certify competency in a set of skills such as EKG reading or CPR

Valid

A good test is valid because the test conditions, behavior and standards are the same as in the objective.

Example

If the objective of the course is for your residents to perform a flexible sigmoidoscopy with little or no discomfort to the patient, given a normal adult presenting for the exam, which is the valid test?

- a. You give a multiple choice test of knowledge of the flex-sig procedure.
- b. You give the resident a simulated patient, the flex-sig apparatus, and test instructions to perform the flex-sig procedure with little or no discomfort to the patient.

The answer is "b." The test conditions, behavior and standards more closely match those of the objective.

Reliable

A good test is reliable because it provides sufficient evidence of the desired knowledge or skill. Would you want to be operated on by a surgeon who had performed the procedure only once successfully? Twice? Three times? You would probably want to make sure the surgeon had done this procedure many times under competent supervision before doing it on his/her own, and before doing it on you!

Reliable testing means that you know the learner really knows his/her stuff, or can really do the procedure...it wasn't just by luck!

Objective

A good test is objective, meaning that two trained evaluators would obtain the same result with the test.

Example

You and a colleague are given the task to conduct a performance evaluation of interviewing skills at the end of the first year of medical school. After the evaluation you discover that all your evaluation results are significantly lower than those of your colleague. Assuming that the students you evaluated were pretty much the same as your colleague's, you probably have to assume that either you were too harsh an observer or your colleague was too lenient an observer.

The solution would be for you and your colleague to standardize your observations by discussing criteria for acceptable performance, observing a limited set of student performances (these could be taped or live), recording and discussing your observations.

Comprehensive

A good test tests on samples of all the course requirements.

Example

You have developed a course in introductory clinical skills that includes the following:

- using a stethoscope to listen to the heart and lungs
- using an otoscope to examine the ear
- using an ophthalmoscope to examine the eye
- using a reflex hammer to test reflexes
- using a blood pressure cuff to check a blood pressure

How many of these skills will you need to "sample" for your test to be comprehensive?

(Answer: all of them)

Differentiating

A good test differentiates between those who know and those who do not know. This means that competent learners score high on the test and incompetent learners score low.

Example

Students at a little-known medical school performed inconsistently on Part 1 of the boards. Some students passed with flying colors, some got by and others failed outright. Students were at a loss to explain this as they had all done well on their final exams for all their basic science courses at the medical school. Unfortunately, they were victims of poorly differentiating tests which did not separate out the poor performers from the solid performers. The tests were too easy to be differentiating.

Expected

A good test tests what is expected. Tell learners the objectives so they can practice what is on the test.

Example

You take a course on sailing. The course objectives include learning how to sail, and the final exam includes the instructor observing your performance in a sailboat. Unfortunately, while you get plenty of "book learning," you and your fellow students each get only 15 minutes at the helm. The expectations for your performance on the test are out of line with the expectations for practice.

Instructive

A good test is instructive because it tells the test-taker what was well done or what is in need of improvement.

Which is more useful to the learner, a red "x" indicating an incorrect response, or a sentence explaining why something is correct or incorrect? Good feedback is one of the qualities of good learner assessment.

Useful

A good test is useful if it can be implemented with available resources.

Objective structured clinical exams, in which simulated patients and trained observers are used to help assess students' clinical skills, can be very valid, objective, instructive, differentiating, and comprehensive. They can also be very expensive, requiring much in terms of money, space, and time. For this reason, many institutions have abandoned the approach.

PART 2: General Standards for Test Items

Just as there are qualities that good tests share, there are also a number of general standards that should be observed no matter what type of test items you write. In this unit you will learn what those general standards are, and why you should use them.

There are nine general standards for test items:

1. There are sufficient test items to cover all important ideas.
2. All test items are on ideas related to instructional objectives.
3. Questions simulate the use of the test content as specified in the objective.
4. All questions are easy to read.
5. There is only one right answer.
6. Don't give the answer away.
7. The answer to one question doesn't affect the answer to another question.
8. Questions are grouped by type.
9. Questions refer to a source if needed.

1. There are sufficient test items to cover all important ideas.

In any well-designed course an important, defined body of knowledge or set of skills comprises the content of instruction. A conscientious teacher will want to learn if students have mastered the course knowledge and skills. If content is important enough to teach and learn, it is important enough to test.

Thus, you as a course designer will need to ensure that, for each idea or skill taught, there is a corresponding test item. As we will see later, the form of the test item will vary according to the kind of knowledge and skill tested.

2. All test items are on ideas related to instructional objectives.

This standard comes straight from the *Secret of Instructional Design* you first learned in September. The secret states that all instructional components (goals, objectives, content, methods and evaluation) must be present and consistent with each other. If you are going to include an item in your test, make sure it was part of the course objectives, and was addressed either through direct instruction, readings, self-study, discussion, or some other method.

3. Questions simulate the use of the test content as specified in the objective

Here is an example of a test item that simulates the use of the content.

Example

Objective: Given a stethoscope and blood pressure cuff, the learner will be able to take a patient's blood pressure within 30 seconds, accurate to within five millimeters of mercury.

Content: Procedure for taking a blood pressure. There are several ways to test learners on this content:

1. Have learners recall and write down the steps for taking a blood pressure.
2. Present the steps for taking a blood pressure, and ask learners to arrange them in their proper order.
3. Have learners take a blood pressure on a simulated patient.

Question: Given the objective, what do you think is the best way of testing this content on the final exam? Answer #3, because it is the closest simulation of the use of the content, as specified in the objective.

Question: If the objective was "Learners will be able to recall the steps for taking a blood pressure," what would be the right test item? Answer #1, because it most closely simulates the use of the content as specified in the objective.

4. All questions are easy to read.

Test takers should not have to guess at the meaning of a test item but only apply what they've learned to answer it. Thus, test items should be grammatically correct and clear with concise, deliberate word choices.

5. There is only one right answer.

Compare these two items:

- 1) Lister is best-known for_____.
- 2) Lister used which chemical agent to introduce antisepsis in surgery?_____

There are several possible answers to item 1:

- a) Introducing antisepsis in surgery.
- b) Using carbolic acid to introduce antisepsis in surgery.
- c) Using heat-sterilized instruments to help prevent infection.

The ambiguity of item 1 invites a variety of responses, some more complete than others. You can avoid this by being very specific in what you are asking for in the test item.

6. Don't give the answer away.

The answer to one question shouldn't give away the answer to another question. This is a favorite for learners who are "test-wise." Other ways of giving away the answer are:

- Making one response of a multiple-choice answer consistently longer than the others. Using qualifiers to make sure a multiple choice answer is correct can lengthen that answer. You should vary the length of correct answers or better yet, try to make all choices of approximately equal length.
- Grammatical cues such as "a" versus "an" at the end of a stem in a multiple choice item. If your test items require completion of a sentence by choosing the correct multiple choice response, make sure there are no grammatical cues that either eliminate incorrect foils or lead to the correct choice.
- Using words such as "always," "never," "sometimes," "all," and "normally," especially in true-false items. They may give a clue as to the correct response. "Never" is rarely "never" and "always" is rarely "always," so an item using these words would probably be false.
- Position: having the correct answer appear as response "b" in multiple choice items more often than in the other positions. This situation arises often when a test is prepared in haste, and the instructor falls into an item writing pattern.

Example

Compare these two items where the first question gives away the second:

- 1) A cardiologist prescribes "Bustaclo43" to control hypertension in a 55-year old male patient with diabetes. What of the following side effects is most likely?
 - a) lower serum glucose
 - b) elevated serum glucose
 - c) lower sodium levels
 - d) elevated sodium levels
- 2) Bustaclo43 is currently used to:

- a) relieve migraine headache pain
- b) control hypertension
- c) manage hypercholesterolemia
- d) alleviate iron deficiency

The test-wise student will recall the clue embedded in question 1 when answering question 2. Therefore, when you write your test items, you want to be sure that your stems do not provide information that can help your learners answer other test items.

7. The answer to one question doesn't affect the answer to another question.

Items that require the test-taker to use information from another answer are unfair in that they can penalize the learner twice or more times for one mistake. Problems like these arise especially in items that require computation. If a number derived in one answer must be "plugged into" a formula in another test item, the consequences for answering the first item incorrectly are multiplied.

8. Questions are grouped by type.

Put all T/F questions together, all multiple choice, all matching, and so on. In this way learners do not have to figure out how to answer each question.

9. Questions refer to a source if needed.

There may be differences of opinion among experts regarding certain issues. For example, the use of PSA may be advocated by some physicians for all men over 45 years old, while others may feel it isn't worth the effort at this stage. Whenever differences of opinion may arise, it is important to include a source for the information.

Example

T___F___ PSA should be given to all men over the age of 45.

T___F___ According to Hartman, PSA should be given to all men over the age of 45.

PART 3: Technical Specifications for Types of Test Items

There are a number of technical specifications that should be observed depending on the type of test items you write. In this unit you will learn what those technical specifications are, and why you should use them. We will cover a variety of test items in this section including questions where the learner recognizes the answer (true/false, multiple choice, matching) and questions where the learner supplies the answer (completion, short answer, essay, production, procedure: criterion checklists).

To get the most out of this section, have at your side a number of test items you've developed for your tests and compare them with the specifications. If they don't "measure up," use the specifications for each type of test item to figure out ways to improve them.

True or False Questions

A true or false question is essentially a statement, called a proposition. The learner judges whether the proposition is true or false.

Example

Vitamin C is useful in the prevention of scurvy.

True or false questions can cover a larger amount of subject matter than any other test item. Every true or false question should follow these technical specifications.

Technical Specifications

1. There should be about an equal number of true and false statements.
2. Both true and false statements should be about equal length.
3. False items should be plausible.

<p>Good Example The MDR of vitamin D for a normal 160 pound male adult is 4,000 units. <i>(assuming the correct response is 400 units)</i></p>	<p>Bad Example The MDR of vitamin D for a normal 160 pound male adult is 4 pounds.</p>
---	---

4. The proposition should be clearly true or false.

<p>Good Example A fracture of the femur protruding through the skin is an example of a compound fracture.</p>	<p>Bad Example A simple fracture of the femur protruding through the skin is an example of a compound fracture. <i>("Simple" and "compound" as technical terms are mutually exclusive.)</i></p>
--	--

5. The proposition should be simple.

<p>Good Example According to Street, 42.3 out of 127 people will tell the truth on a survey.</p>	<p>Bad Example According to Street, 42.3 out of 127 people interviewed at random will proffer credible responses sufficient to satisfy the requirements for veracity underlying reliable, valid survey research.</p>
---	---

6. The proposition should be specific.

Good Example 15% of all American males aged 55 and over suffer from cardiac arrhythmia.	Bad Example About 25 - 35% of Americans suffer from some sort of heart condition.
---	---

7. The proposition should contain one idea.

Good Examples Vitamin C is useful in the prevention of scurvy. Vitamin D is useful in the prevention of scurvy.	Bad Example Vitamin C and Vitamin D are useful in the prevention of scurvy.
--	---

8. The proposition should include sources if needed.

Good Example According to Street, 42.3 out of 127 people will tell the truth on a survey.	Bad Example 42.3 out of 127 people will tell the truth on a survey.
---	---

9. Avoid broad statements. Statements with too broad a scope are usually false.

Good Example High levels of LDL (>180 mg/dl) is a risk factor for heart disease in men over 40.	Bad Example Cholesterol is bad for men over 40.
---	---

10. Avoid trivia.

Good Example The incidence of tuberculosis among all people living in Greenville, NC is .05%	Bad Example The incidence of Crumbler's Syndrome among blue-eyed Caucasians from Greenville, NC is .0000000005%.
--	--

11. Avoid qualifiers. Adjectives that qualify the meaning of a noun, such as "always" and "never," usually make the statement false.

Good Example AIDS is characterized by the co-occurrence of opportunistic infections.	Bad Example AIDS is always accompanied by opportunistic infection.
--	--

12. Avoid indefinites.

Good Example AIDS is characterized by the co-occurrence of opportunistic infections.	Bad Example Some AIDS patients suffer from opportunistic infections.
--	--

13. Avoid negative statements.

Good Example One can use a summative evaluation design for formative evaluation purposes.	Bad Example One should not use a formative evaluation design for summative evaluation purposes.
---	---

Multiple Choice Questions

A multiple choice item consists of a stem, which contains the problem, and a list of suggested responses. The incorrect responses are called "foils" or "distracters."

Example

- Which federal agency is responsible for financing Medicare payments? (stem)
- a. HRSA (foil)
 - b. HCFA (answer)
 - c. HUAC (foil)
 - d. NIH (foil)

Multiple choice questions are some of the most useful test items. You can test everything from factual recall to application of principles to problems. Every multiple choice question should follow these specifications.

Technical Specifications

Stems

1. Most information is given in the stem.

Good Example	Bad Example
Following a second episode of salpingitis, what is the likelihood that a woman is infertile? a. Less than 20%. b. 20 to 30% c. etc.	Following a second episode of salpingitis: a. the likelihood that a woman is infertile is less than 20%. b. the likelihood that a woman is infertile is 20 to 30% c. the likelihood that a woman is infertile is greater than 50% d. etc.

2. Only relevant and realistic information is provided.

Good Examples	Bad Examples
A 68-year-old man is scheduled for a sports physical. (It's possible!) A 48-year-old female with a history of psychiatric hospitalization and alcohol and drug use is confused and non-cooperative. She speaks of experiencing complete detachment from the world around her. The most likely psychiatric diagnosis is:	A 110-year-old man is scheduled for a sports physical. (Not very likely) A 48-year-old female with a history of psychiatric hospitalization, high cholesterol, heart disease, alcohol and drug use, and morbid obesity is confused and non-cooperative. She speaks of experiencing complete detachment from the world around her. The most likely psychiatric diagnosis is:

3. Provide new examples and cases if needed.

Example
A 15-year-old female presents with the following symptoms: 30 pound weight loss in the last two months, advanced tooth decay, blood pressure 90/40, and temperature of 94.5 degrees F. What should you do first? <i>This is a good example if the case has not been previously seen. It is a bad example if the case has been previously seen.</i>

4. The stem should be a whole, positive statement.

Good Example "Schistosomiasis" is also known as:	Bad Example "Schistosomiasis" is never referred to as:
--	--

5. The stem should avoid using grammar or word cues to give away the answer.

Good Example The most common medical prophylaxis for malaria is: a. nivaquine b. privaquine c. aralen d. chloroquine e. quinine	Bad Example The most common medical prophylaxis for malaria is: a. mosquito netting b. privaquine c. avoiding mosquito bites d. chloroquine e. using insecticide <i>(Three foils are non-medical, so they can be ignored by the test-wise learner.)</i>
--	---

Answers and Foils

6. Correct answers and foils should be short.

Good Example Following a second episode of salpingitis, what is the likelihood that a woman is infertile? a. Less than 20%. b. 20 to 30% c. Greater than 50% d. etc.	Bad Example Following a second episode of salpingitis: a. there is less than a 20% likelihood that a woman is infertile. b. the likelihood that a woman is infertile is 20 to 30% c. there is a greater than 50% likelihood that a woman is infertile. d. etc.
--	--

7. There should be only one correct answer.

Good Example Normal body temperature for an adult is: a. 98.6 degrees Fahrenheit b. 35 degrees Celsius c. 96.8 degrees Fahrenheit d. 38 degrees Celsius	Bad Example Normal body temperature for an adult is: a. 98.6 degrees Fahrenheit b. 35 degrees Celsius c. 96.8 degrees Fahrenheit d. 37 degrees Celsius
---	--

8. Grammar should be consistent for answer and foils.

Good Example _____ a day keeps the doctor away. a. A banana b. An apple c. A steak d. A peach e. A good jog	Bad Example An _____ a day keeps the doctor away. a. banana b. apple c. steak d. peach e. good jog
--	---

9. Answer and foils should be presented in some logical order.

<p>Good Example The lowest toxic level of digitoxin for a 180 lb male (specimen obtained 12-24 hours after last dose) is: :</p> <ul style="list-style-type: none"> a. 15 nanogm/ml b. 24 nanogm/ml c. 35 nanogm/ml d. 50 nanogm/ml e. 75 nanogm/ml 	<p>Bad Example The lowest toxic level of digitoxin for a 180 lb male (specimen obtained 12-24 hours after last dose) is:</p> <ul style="list-style-type: none"> a. 75 nanogm/ml b. 15 nanogm/ml c. 50 nanogm/ml d. 35 nanogm/ml e. 24 nanogm/ml
--	---

10. Answer and foils should be mutually exclusive.

<p>Good Example The average weight of an adult male in the United States is:</p> <ul style="list-style-type: none"> a. less than 150 pounds b. more than 200 pounds c. between 150 and 170 pounds d. between 171 and 185 pounds e. between 186 and 200 pounds 	<p>Bad Example The average weight of an adult male in the United States is:</p> <ul style="list-style-type: none"> a. less than 150 pounds b. less than 175 pounds c. more than 185 pounds d. more than 200 pounds
---	---

11. Answer and foils should be about equal length.

<p>Good Example An extrovert is:</p> <ul style="list-style-type: none"> a. a person who compulsively steals green objects. b. an introvert with a Weschler I.Q. score above 125. c. a person whose interest is turned to the external world. d. a person with a neurotic need for attention. 	<p>Bad Example An extrovert is:</p> <ul style="list-style-type: none"> a. a person who steals. b. a bright introvert. c. a person whose interest is turned to the external world. d. an obsessive capitalist.
---	--

12. Avoid "none of the above" and "all of the above."

<p>Good Example What is the number obtained when the circumference of a circle is divided by its diameter?</p> <ul style="list-style-type: none"> a. 1.61 b. 2.36 c. 3.14 d. 4.32 	<p>Bad Example What is the number obtained when the circumference of a circle is divided by its diameter?</p> <ul style="list-style-type: none"> a. 1.61 b. 2.36 c. 3.14 d. none of the above (<i>"c" and "d" are correct</i>)
--	---

13. Answer and foils should be similar in difficulty.

<p>Good Example Which federal agency is responsible for financing Medicare payments? a. HRSA b. HCFA c. HUAC d. NIH</p>	<p>Bad Example Which federal agency is responsible for financing Medicare payments? a. Department of Agriculture b. HCFA c. FBI d. Supreme Court</p>
--	---

Matching Items Questions

A matching item question is one that requires the test taker to match an item in one column with an item from a second column. In general, the items that have a blank space next to them are called the "questions" and the items that you choose from to fill in the blank are called the "answers."

Example

Match the hypertension treatment drug with its class. You may use a class more than once.

Drug:

- ___ 1. Propranolol
- ___ 2. Hydrochlorothizide
- ___ 3. Captopril
- ___ 4. Enalapril
- ___ 5. Atenolol

Class:

- a) ACE inhibitors
- b) ACE II inhibitors
- c) alpha antagonist
- d) beta blocker
- e) calcium channel blocker
- f) loop diuretic
- g) thiazid diuretic

Matching items questions are an efficient way of testing simple factual recall such as knowledge of definitions, associations between items and their functions, or examples and their classes. Every matching question should follow these specifications.

Technical Specifications

1. Instructions indicate the basis for matching.

<p>Good Example "Write the letter of the item on the right in the space next to the item on the left. Any item on the right can be used once, more than once, or not at all."</p>	<p>Bad Example "Match the item on the right with the item on the left."</p>
--	--

2. Questions and responses are all from the same category.

<p>Good Example ___ 1. Pasteur ___ 2. Boyle ___ 3. Curie ___ 4. Koch ___ 5. Krebs</p>	<p>Bad Example ___ 1. 1876 ___ 2. Koch ___ 3. pasteurization ___ 4. allopathic ___ 5. first surgical procedure using anesthetic</p>
--	--

3. Questions and responses include 5 to 8 items.
4. There should not be an equal number of test items; 2 to 3 more possible responses should be available.

Good Example

___1. Boyle	a. stated evidence required to establish etiology
___2. Curie	b. founded microbiology
___3. Koch	c. developed antiseptic surgery
___4. Krebs	d. discovered element of atomic number 96
___5. Pasteur	e. distinguished between chemical elements and compounds
	f. shared discovery of citric acid cycle
	g. founded descriptive psychiatry

Bad Example

___1. Boyle	a. stated evidence required to establish etiology
___2. Curie	b. founded microbiology
___3. Koch	c. discovered element of atomic number 96
___4. Krebs	d. distinguished between chemical elements and compounds
___5. Pasteur	e. shared discovery of citric acid cycle

5. Questions and responses should be grammatically consistent; they should not contain any grammatical cues.

Good Example

___1. Boyle	a. stated evidence required to establish etiology
___2. Curie	b. founded microbiology
___3. Koch	c. developed antiseptic surgery
___4. Krebs	d. discovered element of atomic number 96
___5. Pasteur	e. distinguished between chemical elements and compounds
	f. shared discovery of citric acid cycle
	g. founded descriptive psychiatry

Bad Example

___1. Boyle	a. stated evidence required to establish etiology
___2. Curie	b. founded microbiology
___3. Koch	c. developed antiseptic surgery
___4. Krebs	d. woman who discovered element of atomic number 96
___5. Pasteur	e. first to distinguish between chemical elements and compounds
	f. shared discovery of citric acid cycle
	g. founder of descriptive psychiatry

6. Questions and responses should all be on the same page.

7. Responses are short phrases.

Good Example	Bad Example
<p>___1.pyruvate kinase a. Deficiency results in hemolytic anemia</p>	<p>___1 pyruvate kinase a. Deficiency of the enzyme, an autosomal recessive trait, results in hemolytic anemia</p>

8. All the responses are plausible for the questions.

Good Example

<p>___1. creatine kinase</p> <p>___2. pyruvate kinase</p>	<p>a. Deficiency results in hemolytic anemia</p> <p>b. Occurs as three isoenzymes</p> <p>c. Initiates digestion of dietary fats</p>
---	---

Bad Example

<p>___1. creatine kinase</p> <p>___2. pyruvate kinase</p> <p>___3. pyruvate kinas</p> <p>___4. creatine kinase</p> <p>___5. pyruvate kinase</p>	<p>a. Deficiency results in hemolytic anemia</p> <p>b. Genus of gram positive bacterium</p> <p>c. Initiates digestion of dietary fats</p>
---	---

9. Responses are listed to the right of the questions.

10. Responses are in a systematic order: alphabetic for lists of names; numeric for dates and numbers.

Good Example

<p>___1. Boyle</p> <p>___2. Curie</p> <p>___3. Koch</p> <p>___4. Krebs</p> <p>___5. Pasteur</p>	<p>a. Stated evidence required to establish etiology</p> <p>b. Founded microbiology</p> <p>c. Developed antiseptic surgery</p> <p>d. Discovered element of atomic number 96</p> <p>e. First to distinguish between chemical elements and compounds</p> <p>f. Shared discovery of citric acid cycle</p> <p>g. Founder of descriptive psychiatry</p>
---	--

Bad Example

<p>___1. Koch</p> <p>___2. Pasteur</p> <p>___3. Boyle</p> <p>___4. Krebs</p> <p>___5. Curie</p>	<p>a. Stated evidence required to establish etiology</p> <p>b. Founded microbiology</p> <p>c. Developed antiseptic surgery</p> <p>d. Discovered element of atomic number 96</p> <p>e. First to distinguish between chemical elements and compounds</p> <p>f. Shared discovery of citric acid cycle</p> <p>g. Founder of descriptive psychiatry</p>
---	--

Completion Questions

A completion item is a form of short answer question in which the learner completes a sentence by supplying a key word or phrase. A completion item is comprised of two parts, the "cue" and the blank.

Example

The most abundant protein in muscle is _____. (*Answer: myosin*)

Completion questions are the simplest types of test items in which the learner is required to supply the correct answer, rather than to choose the correct answer. As such, it requires a higher level of learning – recall learning – rather than simple recognition. Every completion question should follow these specifications.

Technical Specifications

1. The sentence describes the blank(s).

Good Example

The MDR of vitamin D for a normal 160 pound male adult is _____. (*Answer: 400 units*)

2. Blank(s) are at the end of the sentence. The point of the item can be lost if the blank occurs earlier in the item. Also, learners may have to return to the beginning of the item to recall what is being asked of them.

Good Example

A chronic disease caused by a variety of fungi including *Nocardia brasiliensis* affecting the foot, hands, legs or other parts including the internal organs is _____.
(*Answer: maduromycosis*)

Bad Example

_____ is a chronic disease caused by a variety of fungi including *Nocardia brasiliensis* affecting the foot, hands, legs, or other parts including the internal organs.

3. Blank(s) should be equal in length. Blanks of unequal length may be misread as an additional cue to learners. All blanks should be of sufficient length to accommodate any correct answer for all completion questions.
4. There should be few blanks per question.

Good Example

The ratio of the opposite side to the hypotenuse is called the _____. (*Answer: sine*)

Bad Example

The ratio of the _____ to the _____ is called the _____.

(*There are multiple possibilities for answers: percentage of red cells / percentage of hemoglobin = cell color ratio; carbon dioxide output / oxygen uptake = expiratory exchange ratio.*)

5. Each blank should be worth one point. Assigning differential point values to different blanks is an invitation to charges of unfairness in grading.

Short Answer Questions

A short answer question is a complete question that requires the learner to supply the correct answer. The answer should be brief.

Example

Which managed care organizations have their home base in Lansing?

Short answer questions are another type of question where the learner must supply the answer rather than recognize it from a list of choices. It differs from its close relative, the completion question, in that it poses a question to be answered, rather than a blank to be completed. It differs from the essay question according to the length of its response, which should be brief and specific. All short answer questions should follow these specifications.

Technical Specifications

1. State a direct, full question.

<p>Good Example In what city was the first surgical procedure using anesthetic conducted?</p>	<p>Bad Example City of first surgical procedure using anesthetic.</p>
--	--

2. The question should call for a specific response of a number, symbol, word or phrase.

<p>Good Example What is the value of pi to 4 decimal places?</p>	<p>Bad Example What is the value of pi?</p>
---	--

Essay Questions

An essay question calls for an extended response from the learner. The response can be extended, with virtually no restrictions on the answer, or it can be restricted according to length.

Example

Describe the three most common methods of STD prevention and compare them with respect to effectiveness. Your response should be limited to five paragraphs and should use Emerson's criteria for effective STD prophylaxis.

Essay questions allow the learner maximum freedom to respond. Higher order mental processes can be tested using essay questions such as description, comparison, evaluation and prediction. All essay questions should follow these specifications.

Technical Specifications

1. Essay question is used for an appropriate objective.

<p><i>Objective: Given a written case presentation of a patient presenting with chest pain, the learner will be able to write a differential diagnosis for possible causes that account for the presenting symptoms, providing rationales for each item in the differential.</i></p>	
<p>Good Example Write a differential diagnosis for the patient in the case presentation, giving at least four conditions that would account for his presenting symptoms, Include a rationale for each condition in your differential that will account for at least three of the symptoms, lab findings or physical examination findings.</p>	<p>Bad Example List possible diagnoses for the following case presentation.</p>

2. Establish a framework to guide the student. Lack of a framework delimiting the universe of possible responses to that required by the objectives of instruction can lead to difficulty on the part of learners in determining how to answer a given question. It can also lead to difficulty in grading the response reliably; a variety of answers may be possible depending upon interpretation. To establish a framework, limit the area covered by the question. In the example, a specific type of hypertension is included in the good example, whereas the bad example refers to "hypertension" generically.

Use descriptive words such as *define*, *outline*, *classify*, *summarize*. Avoid the use of the word *discuss*. In the good example, the learner is directed to "describe the effects of (condition) on (organ)," an instruction that limits the response quite specifically when compared with the bad example.

<p>Good Example Describe the effects of splenoportal hyper-tension on the liver. Limit your discussion to one-half page.</p>	<p>Bad Example Discuss the effects of hypertension.</p>
---	--

3. The question is realistic in terms of difficulty, time allotted, and complexity.

<p>Good Example Describe the effect of the psychoactive agent tetrahydracannabinol on the healthy central nervous system. Limit your response to a maximum of one-half page. You will have 15 minutes to answer this question.</p>	<p>Bad Example Describe the effects on the central nervous system of all classes of psychoactive drugs. You have 10 minutes to complete this item.</p>
---	---

4. A model response is prepared for each question. Preparation of model responses will result in more reliable and objective scoring of essay questions. Preparation of model answers as questions are being written will help reduce question ambiguity, unrealistic expectations and other problems.
5. Always tell the time allotted and point value of each question. These indications on a test will give students a basis for pacing themselves.

Production Questions

A production question is a test item that requires the learner to produce a specific written product such as a plan or analysis, according to criteria included or referred to in the question.

Example

Write a description of a legitimate research project that contains an ethical dilemma as defined by Stroke. Your answer should include a description of the research question, the research methods, and the dilemma posed by the project. (15 points)

Production questions are a variation of essay questions, but are very precisely written so that the learner will "produce" a written item according to the instructions. Examples of written items could be plans, analyses or comparisons. All production questions should follow these specifications.

Technical Specifications

1. Explain the directions carefully and specifically.

<p>Good Example Write an objective that includes all four parts according to Mager.</p>	<p>Bad Example Write an objective.</p>
--	---

2. Tell exactly what to do: delimit scope of response, use clear action verbs, and subdivide if necessary.

<p>Good Examples Write a lesson plan that includes all elements of a lesson according to the course checklist. <i>(Has a clear action verb.)</i></p> <p>Write an exercise and diet plan for a patient who wishes to build strength, reduce weight by five pounds in one month, and decrease blood pressure. Use all or part of the following framework. Include only those parts of the plan that will help your patient reach his/her goals</p> <p><u>Patient information:</u> sex: male height: 5 ft 10 in weight: 210 lbs bp (resting): 175/90</p> <p><u>Response framework:</u> Cardiovascular conditioning: Strength: Flexibility: Diet: <i>(Scope of response clearly indicated. Answer is subdivided into desired categories.)</i></p>	<p>Bad Examples Give an example of a lesson plan. <i>(verb not clear)</i></p> <p>Write an exercise and diet plan for a patient with essential hypertension. <i>(Scope of response unclear. Answer is not subdivided into desired categories.)</i></p>
--	--

3. Mention standard.

<p>Good Example Write an example of a complete prescription for a Class 3 narcotic according to the regulations of the State of Michigan governing such prescriptions.</p>	<p>Bad Example Write an example of a complete prescription for a Class 3 narcotic.</p>
---	---

4. Prepare a checklist and model response. By preparing a checklist of items to be included in a response, and a model response, you are creating criteria for correct responses. You can then refine the wording of the question such that learners will clearly understand what is expected of them.

Example

Production Test Question: "Write an objective that includes all four parts according to Mager."

Checklist:

- ___ conditions
- ___ behavior
- ___ standards
- ___ lower limits

Model response:

Given a written case study of a patient presenting with chest pain (*condition*), the learner will write a differential diagnosis of possible causes of the symptoms (*behavior*), including a written rationale for each possible cause (*standard*), according to Emerson (*standard*), with no omissions (*lower limit*).

5. Always state the time allotted and the point value for each question. These indications on a test will give students a basis for pacing themselves.

Procedures: Criterion Checklists

The content and examples in this section are adapted from "How to Use - and Create - Criterion Checklists" by S. Yelon, *Performance and Instruction Journal*, April, 1984.

Procedures, also referred to as psychomotor skills, are best tested using "criterion checklists." A criterion checklist states both:

- a) the steps in a procedure that must be included in an acceptable performance
- b) the quality or degree of excellence with which the steps must be performed

Every criterion checklist should follow these specifications.

Technical Specifications

1. State the steps of the performance or the components of the product. Ask yourself, "How does a successful person do this skill? What does a good finished product look like?"

<p>Example #1 <u>Procedure</u></p> <p>Taking a blood pressure</p> <p><u>Checklist</u></p> <p>Step 1: Get equipment together: stethoscope, alcohol swabs, blood pressure cuff.</p> <p>Step 2: Position patient</p> <p>Step 3: Explain procedure to patient</p> <p>Step 4: Wrap cuff, etc.</p>	<p>Example #2 <u>Product</u></p> <p>Essay defending a proposition</p> <p><u>Checklist</u></p> <p>Opening paragraph: a. state proposition b. paragraph has attention-getter c. ends with transition to next paragraph. d. etc.</p> <p>Supporting paragraphs: a. each paragraph support/explains proposition. b. transitions to/from each paragraph c. etc.</p> <p>Concluding paragraph: a. all ideas tied together b. etc.</p> <p>Grammar and spelling are correct.</p>
---	---

2. State the criteria for each step and each component. Ask yourself what makes these steps or attributes acceptable or of high quality.

<p>Example #1 <u>Procedure</u></p> <p>Taking a blood pressure</p> <p><u>Checklist</u></p> <p>Step 1: Get equipment together: stethoscope, alcohol swabs, blood pressure cuff. (chooses appropriate size cuff for patient)</p> <p>Step 2: Position patient (comfortable, lying or sitting, arm supported and palm upward)</p> <p>Step 3: Explain procedure to patient (check for understanding)</p> <p>Step 4: Wrap cuff, etc. (no clothing between cuff and arm, cuff without wrinkles, snug but still can fit a finger between cuff and arm)</p>	<p>Example #2 <u>Product</u></p> <p>Essay defending a proposition</p> <p><u>Checklist</u></p> <p>Opening paragraph:</p> <ul style="list-style-type: none"> a. state proposition (concrete terms, complete thought) b. paragraph has attention-getter c. ends with transition to next paragraph. d. etc. <p>Supporting paragraphs:</p> <ul style="list-style-type: none"> a. each paragraph support/explains proposition. b. transitions to/from each paragraph c. etc. <p>Concluding paragraph:</p> <ul style="list-style-type: none"> a. all ideas tied together b. etc. <p>Grammar and spelling are correct (spell checked, complete sentences, correct word choice).</p>
--	---

3. Assemble a draft checklist and establish cutoff standards. Include a **checkpoint** for presence or absence of a step in procedure or component of a product and any necessary qualities.

<p>Example #1 <u>Procedure</u></p> <p style="padding-left: 40px;">Taking a blood pressure</p> <p><u>Checklist</u></p> <p><input checked="" type="checkbox"/> 1. Get equipment together: stethoscope, alcohol swabs, blood pressure cuff.</p> <p><input type="checkbox"/> 2. Choose appropriate size cuff for patient.</p> <p><input type="checkbox"/> 3. Position patient :</p> <p style="padding-left: 40px;"><input type="checkbox"/> comfortable, lying or sitting</p> <p style="padding-left: 40px;"><input type="checkbox"/> arm supported and palm upward</p> <p><input type="checkbox"/> 4. Explain procedure to patient.</p> <p><input type="checkbox"/> 5. Check for understanding.</p> <p><input type="checkbox"/> 6. Wrap cuff.</p> <p style="padding-left: 40px;"><input type="checkbox"/> no clothing between cuff and arm,</p> <p style="padding-left: 40px;"><input type="checkbox"/> cuff without wrinkles,</p> <p style="padding-left: 40px;"><input type="checkbox"/> snug but still can fit finger between cuff and arm</p> <p><input type="checkbox"/> 7. etc.</p>	<p>Example #2 <u>Product</u></p> <p style="padding-left: 40px;">Essay defending a proposition</p> <p><u>Checklist</u></p> <p>Opening paragraph</p> <p style="padding-left: 40px;"><input type="checkbox"/> a. state proposition</p> <p style="padding-left: 80px;"><input type="checkbox"/> 1. in concrete terms,</p> <p style="padding-left: 80px;"><input type="checkbox"/> 2. as complete thought</p> <p style="padding-left: 40px;"><input type="checkbox"/> b. paragraph has attention getter</p> <p style="padding-left: 40px;"><input type="checkbox"/> c. ends with transition to next paragraph.</p> <p style="padding-left: 40px;"><input type="checkbox"/> d. paragraph includes statement that explains organization of essay</p> <p>Supporting paragraphs</p> <p style="padding-left: 40px;"><input type="checkbox"/> a. each paragraph support/explains proposition.</p> <p style="padding-left: 40px;"><input type="checkbox"/> b. transitions to/from each paragraph</p> <p>Concluding paragraph</p> <p style="padding-left: 40px;"><input type="checkbox"/> a. all ideas tied together</p> <p style="padding-left: 40px;"><input type="checkbox"/> b. conclusion reinforces main point</p> <p>Grammar and spelling are correct</p> <p style="padding-left: 40px;"><input type="checkbox"/> a. spell checked</p> <p style="padding-left: 40px;"><input type="checkbox"/> b. complete sentences</p> <p style="padding-left: 40px;"><input type="checkbox"/> c. correct word choice</p>
---	--

4. Add points or weights and establish cutoff standards. You may want to assign points or weights to the steps in a procedure or qualities in a product. One system is to determine which steps or qualities are essential—the absence of one or more would result in failure of the test—and those that are optional—they contribute to the quality of a performance or product, but their absence of would not result in failure.
5. Use the checklist to evaluate a performance or a product and then revise the checklist. It is beneficial to try using the checklist to rate a performance or product before using it in a testing situation. You may find missing elements or ambiguous criteria that need revising.

References

1. Marrelli, A. *Writing multiple-choice test items*. Performance and Instruction, 34(8); 24-29, September, 1995
2. Mehrens W and Lehmann I. *Measurement and Evaluation in Education and Psychology* (3rd ed). New York: Holt, Rinehart and Winston, 1984.
3. Yelon, S. *How to use - and create - criterion checklists*. Performance and Instruction, 1-4, April 1984
4. Yelon, S. *Assessing Knowledge & Skill: Good Tests, Papers & Projects* (unpublished)

APPENDIX: TEST ITEM WORKSHEETS

Use these worksheets to guide you in constructing your test items. They include a summary of the technical specifications or steps involved, and provide a place to state the objective and to write the test item, all on the same page.

True/False Questions	31
Multiple Choice Questions	33
Matching Questions	35
Completion Questions.....	37
Short Answer Questions	39
Essay Questions	41
Production Questions	43
Procedures: Criterion Checklists.....	45

True/False Questions

Objective(s):

Technical Specifications:

1. About equal number of true and false statements
2. About equal length for both true and false statements
3. False items should be plausible.

Proposition stated is:

4. clearly true or false.
5. simple.
6. specific.
7. contains one idea.
8. includes sources if needed.

5 pitfalls to avoid:

1. Broad statements
2. Trivia
3. Qualifiers
4. Indefinites
5. Negative statements

Draft your test items in the space below:

1.

2.

3.

Multiple Choice Questions

Objective(s):

Technical Specifications:

5 for stems:

1. Most information is given in the stem.
2. Only relevant and realistic information is provided.
3. New examples and cases if needed.
4. Whole positive stem
5. No grammar or word cues

8 for answers and foils:

1. Correct answer and foils should be short.
2. There should be only one correct answer.
3. Grammar should be consistent for answer and foils.
4. Answer and foils should be presented in some logical order.
5. Answer and foils should be mutually exclusive.
6. Answer and foils should be about equal length.
7. Avoid "none of the above" and "all of the above."
8. Answer and foils should be similar for difficulty.

Draft your test items in the space below:

1.

2.

3.

Matching Questions

Objective(s):

Technical Specifications:

1. Instructions indicate the basis for matching.

Questions and responses:

2. Are all of the same category.

3. 5 - 8 items.

4. Not an equal number of items - 2-3 more possible responses

5. Grammatically consistent

6. All on the same page

Responses:

7. Are short, i.e., single word or phrases.

8. Are all plausible for questions.

9. Are listed to the right of questions.

10. In some systematic order.

Draft your test items in the space below:

1.

2.

3.

Completion Questions

Objective(s):

Technical Specifications

Sentence describes blank(s).

Blank(s) at the end of sentence.

Blanks are equal length.

Few blanks per question.

Each blank is worth one point.

Draft your test items in the space below:

1.

2.

3.

4.

5.

Short Answer Questions

Objective(s):

Technical Specifications:

1. State a direct full question.
2. Call for a specific response:
number, symbol, word, phrases

Draft your test items in the space below. Include model answers:

- 1.
- 2.
- 3.
- 4.

Essay Questions

Objective(s):

Technical Specifications:

1. Essay question is used for an appropriate objective.
2. Establish a framework to guide student.
3. Realistic in terms of difficulty/time allotted/complexity.
4. Model response prepared for each question.
5. Tell time for question and point value.

Draft your test items in the space below. Include model answers:

- 1.
- 2.

Production Questions

Objective(s):

Technical Specifications: <ol style="list-style-type: none">1. Give the information to use.2. Tell exactly what to do: delimit scope of response, use clear action verbs, subdivide if necessary.3. Mention standard.4. Prepare checklist and model response.5. Tell time for question and point value.	Draft your test items in the space below. Include model answers: <ol style="list-style-type: none">1.2.
--	---

Procedures: Criterion Checklists

<p>Objective(s):</p>

5 steps to creating a criterion checklist:	Steps in performance or components of product:	Criteria for each step or component:
<ol style="list-style-type: none"> 1. State the steps of the performance or the components of the product. 2. State the criteria for each step and each component. 3. Assemble a draft checklist and establish cutoff standards. 4. Add points or weights and establish cutoff standards. 5. Use the checklist to evaluate a performance or a product and then revise the checklist. 		
<p>5 steps to creating a criterion</p>		

<p>checklist:</p> <ol style="list-style-type: none">1. State the steps of the performance or the components of the product.2. State the criteria for each step and each component.3. Assemble a draft checklist and establish cutoff standards.4. Add points or weights and establish cutoff standards.5. Use the checklist to evaluate a performance or a product and then revise the checklist.		
--	--	--