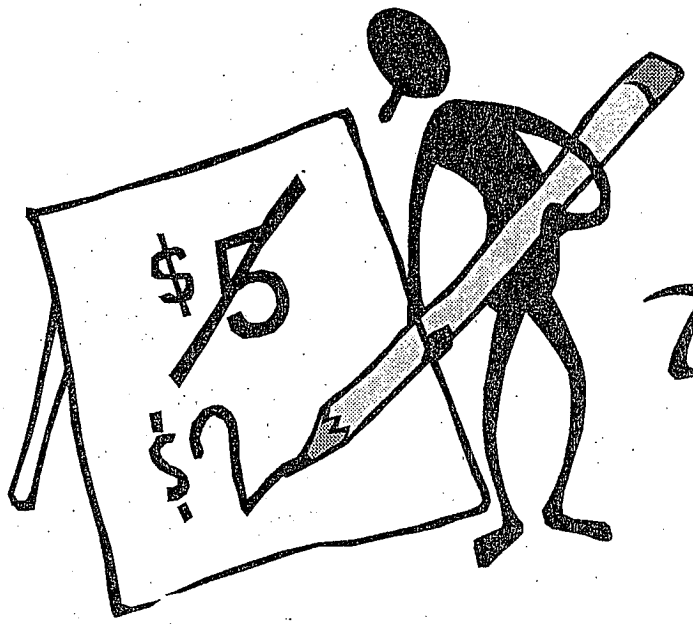

Designing Effective Paper and Pencil Tests

Japan Educators' Days

October, 2000

Laurel Eisinger

Possible Links Between Achievement Targets and Assessment Methods

Target to be Assessed	Assessment Method				
	Knowledge Mastery	**Reasoning Proficiency**	**Skills**	**Ability to Create Products**	**Dispositions**
Selected Response	Multiple choice, true/false, matching, and fill-in can sample mastery of elements of knowledge	Can assess understanding of basic patterns of reasoning	Can assess mastery of the knowledge prerequisites to skillful performance—but cannot rely on these to tap the skill itself	Can assess mastery of knowledge prerequisite to the ability to create quality products—but cannot use these to assess the quality of products themselves	Selected response questionnaire items can tap student feelings
Essay	Can tap understanding of relationships among elements of knowledge	Written descriptions of complex problem solutions can provide a window into reasoning proficiency			Open-ended questionnaire items can probe dispositions

Test Specification Chart

Content	Know and Understand	Comparative Reasoning	Classification Reasoning	Total
Alternative forms of government	9 questions	5	1	15
Structure of U.S. Government	4	5	1	10
Rights and Responsibilities of Citizens	7	5	3	15
Total	20	15	5	40

Examples of Common Sources of Bias in Classroom Assessment

1. Problems common to all methods:

A. Potential problems that can occur within the student:

Lack of reading proficiency
Language barriers
Emotional upset
Poor health
Physical handicap
Peer pressure to mislead assessor
Lack of motivation at time of assessment
Lack of testwiseness (understanding how to take tests)
Lack of personal confidence leading to evaluation anxiety

B. Possible problems that can occur within the assessment context:

Noise distractions
Poor lighting
Discomfort
Lack of rapport with assessor
Cultural insensitivity in assessor or assessment
Lack of proper equipment

C. Examples of problems that arise from the assessment itself (regardless of method)

Directions missing or vague
Poorly worded questions
Poor reproduction reduces readability

2. Problems unique to each format:

A. Possible problems with multiple choice tests:

More than one correct response
Incorrect scoring key
Incorrect bubbling on answer sheet
Clues to the answer in the item or in other items

B. Problems with essay assessments

Students lack writing skill
No scoring criteria
Inappropriate scoring criteria
Evaluator untrained in applying scoring criteria
Bias due to stereotypic thinking or knowledge of prior performance
Insufficient time or patience to read and score carefully

C. Potential problems with performance assessment

No scoring criteria
Inappropriate scoring criteria
Evaluator untrained in applying scoring criteria
Bias due to stereotypic thinking or knowledge of prior performance
Insufficient time or patience to observe and score carefully

D. Possible difficulties when using personal communication

Insufficient sample per student
Inaccurate record keeping
Distortions in memory of performance
Bias due to stereotypic thinking or knowledge of prior performance

COMPARISON OF PAPER AND PENCIL ITEM TYPES

Item Type	Used When	Advantage	Limitations
Multiple Choice	There is only one right answer. There are several plausible alternatives to the correct answer.	Can measure a variety of objectives. Easy to score. Can cover lots of material efficiently.	Can't measure extended thinking or performance. Not good for assessing the process by which answers are obtained.
True/False	A large domain of content is to be tested, requiring the use of many test items.	Can ask many questions in a short time. Easy to score.	Can't measure extended thinking or performance Not good for assessing the process by which answers are obtained. Can be trivial or misleading if not written carefully.
Matching	There are many related thoughts or facts; you want to measure association of information.	Can cover lots of material efficiently. Easy to score.	Can't measure extended thinking. Assess identification of an answer rather than production.
Completion	A clear, short answer is required.	Assessing production of a response. Reduces the possibility of guessing. Can cover lots of material efficiently.	Can't measure extended thinking or performance. Takes longer to score.
Essay/Written	Can measure extended thinking. Can assesse student ability to organize thoughts and compose an appropriate response.	Assesses production of a response, not just identification of a response. Chance plays little or no part in adequate responding. Can be used to assess the process by which an answer is obtained.	Harder to score reliably. Harder to come up with criteria for scoring. Provides a limited sample of knowledge areas. Inability to write can interfere with ability to show understanding.

Guidelines for Test Items

(Taken from Mr. Rick Stiggins, Student Involved Classroom Assessment)

General Guidelines for all formats

- ☐ Items clearly written and focused
- ☐ Question posed
- ☐ Lowest possible reading level
- ☐ Irrelevant clues eliminated
- ☐ Items reviewed by colleague
- ☐ Scoring key double checked

Guidelines for multiple-choice items

- ☐ Item stem poses a direct question
- ☐ Repetition eliminated from response options
- ☐ One best or correct answer
- ☐ Response options are brief and parallel
- ☐ Number of response options offered fits item context

Guidelines for true/false items

- ☐ Statement is entirely true or false as presented

Guidelines for matching exercises

- ☐ Clear directions given
- ☐ List of items to be matched is brief
- ☐ List consists of homogeneous entries
- ☐ Response options are brief and parallel
- ☐ Extra response options offered

Guidelines for fill-in items

- ☐ A direct question is posed
- ☐ One blank is needed to respond
- ☐ Length of blank is not a clue

GUIDELINES FOR WRITING SOUND PAPER AND PENCIL TEST ITEMS

Consult *Student-Centered Classroom Assessment, 2nd Ed.,* Chapter 8 & 9 for examples and greater detail on guidelines.

General Guidelines (apply to all item formats)

1. Keep wording simple and focused:
    a. Follow rules of grammar
    b. Eliminate superfluous material
    c. Come to the point
    d. Test mastery of material, not ability to figure out what you're asking
    e. Aim for lowest possible reading level
2. Ask a question (minimize use of incomplete statements).
3. Avoid providing clues within and between items:
    a. Avoid use of specific determiners such as "always" and "never"
    b. Avoid grammatical clues (e.g., "a" and "an", present and past tense, singular and plural, etc.)
    c. Avoid information clues (e.g., the stem of one item gives away the answer to another)
4. The correct answer should not be obvious to those who have not mastered the materials tested.
5. Highlight critical words such as MOST, LEAST, EXCEPT, and NOT, because they are easily overlooked.

Multiple Choice

1. State the whole question in the test item stem.
2. Keep responses brief and parallel in:
    a. Length
    b. Use of specific determiners
    c. Grammatical construction
    d. Level of generality
3. Avoid repetition of material in each of the response options.
4. Limit use of "all or none of the above".

5. Be sure there is only one correct or best answer (best answer items should be worded to ask for BEST answer).

6. OK to vary number of response options across items within the same test.

True/False

Make them entirely true or false as stated.

Matching

1. Include only homogeneous items. Do not mix dates, events, names, etc. in a single exercise.

2. Maximum length is 10, shorter is better.

3. Provide more responses than there are things to be matched.

4. Provide directions for the match to be made. Indicate if a response can be used more than once or if an item has more than one match.

Completion or fill-in

1. One blank per item (put blank toward end).

2. Length of blank should not be a clue.

Essay

Exercises

1. Avoid general, all encompassing questions.

2. Goal: test structure of knowledge and/or thinking.

3. Point direction to appropriate response in the essay exercise.

4. Give points and/or time allocation.

Scoring

5. Outline acceptable response in advance.

6. Set policy regarding non-achievement factors, i.e., writing.

7. Score in the blind, if possible.

8. Score all responses to one exercise at a time (faster!).

9. Score holistically or analytically, but with criteria clearly established.

Formatting Test Items

1. Be consistent in the presentation of an item-type.

2. Keep all parts of a test question on one page.

3. Avoid crowding too many questions on one page.

Arranging Test Items

1. Group the questions on the test by item-type--that is, put all completion questions together, all multiple choice, all true/false, and so on.

2. Begin with some very easy questions to give students a chance to "break into" the test.

Writing Directions

1. Write clear, explicit directions for each type of item.

2. State the point value of each item-type.

3. Indicate how the answer should be expressed by the student. For example, should the word TRUE or FALSE be written or simply T of F? Should numbers be rounded to the nearest tenth? Should units, such as months, meters, or grams be included in the answer?

Producing Tests

1. Avoid writing tests on the chalkboard.

2. Type tests for duplication--unless your handwriting is very clear and neat.

3. Proofread the test carefully.

4. Duplicate clear, readable copies.

5. Ask a colleague to review or to take important tests.

INTEGRATING ASSESSMENT AND INSTRUCTION--10 SPECIFIC IDEAS

1. Develop a table of test specifications in advance of teaching the unit, so as to define a clear vision of the instructional and assessment target.

2. Share the objectives and table of test specification with each student at the outset of the unit, so as to provide them with a clear vision of their target.

3. Involve students in the development of the objectives and table of specification.

4. Review and modify the vision of targets on your own or along with students as the unit unfolds.

5. Develop assessment exercises along the way--a few each day as the unit unfolds, so as to assure a close match between daily instructional priorities, and so the test will be done and ready to go when it is needed at the end.

6. Involve students in writing some of the items, so as to help them zero in on the targets.

7. Have students analyze the questions on the test after they take it to see if they agree that they really do reflect the agreed upon target.

8. Have students predict their performance on the test based on their knowledge of the agreed upon target.

9. Have students analyze their own responses in terms of the various components of the target to identify their own strengths and weaknesses.

10. Keep tables of test specifications and lists of unit objectives in one file and save the associated test items for reuse when you reteach that unit again in the future. Note cells of the table or individual objectives where students did not do so well, so you can revise instruction next time.