

Objectivity in Designing EFL Testing Items

Ali ZENAGUI Maitre-assistant, Université de Mascara, Algérie.

Abstract

Chances are that some of our strongest childhood and adolescent memories include test taking in school. Language testing, as a measure of evaluation system, is an important part of ELT, as it is strongly held that testing is an effective means of checking syllabus implementation and evaluating teaching quality. Still, getting such valuable data about the teaching/learning process depends largely on the effectiveness of the suggested tests for this purpose. The present paper discusses the importance of selecting objective test-items, and how to construct them by analyzing some of the frequently used test formats in language teaching. Objective test items include items with the following formats: true-false, matching, multiple-choice, and completion or short answer.

Introduction

Tests are necessary and vital to the educational process, as they may represent irrefutable evidence that learning has occurred. They are generally seen as tools that can effectively enhance the learning process, but only if they are well designed and properly used. However, tools can be used and abused, and we are acutely aware that both (poorly and well-designed) tools in the hands of ill-trained or inexperienced users can be dangerous. Therefore, writing objective test items is one of the crucial parameters teachers should take into consideration when designing tests.

The present paper discusses objective test items and to construct them. Items dealt with, here, include: True/False Statements, Matching Items, Multiple Choice Questions

Once we have reached the item-writing stage of test construction, we will have to choose a format or combination of formats to use for our test. Although the choice can be somewhat arbitrary at times, this is not always the case. Often our decision has already been made for us, or more correctly, we may have at least partially made the decision when we wrote the objective or objectives. In many instances, however, we will have a choice among several item formats. For example, let's consider the following objectives and item formats:

OBJECTIVE 1: Given a story, the student can recognize the meaning of all new words.

TEST ITEM: Circle the letter that represents the correct meaning of the following words.

- | | |
|--------------|-----------|
| 1. intention | 2. crisis |
| a. desire | a. |
| feeling | |
| b. need | b. |
| message | |
| c. direction | c. |
| pending | |
| d. command | d. |
| critical | |

OBJECTIVE 2: The student can associate the characteristics of leading characters with their names.

TEST ITEM: The first column below is a list of the names of the main characters in *Huckleberry Finn*. Descriptions of the main characters are listed in the second column. In the space provided, write the letter of the description that matches each character.

| <i>Characters</i> | <i>Descriptions</i> |
|----------------------|----------------------------------|
| _____ 1. Tom | a. Cruel |
| _____ 2. Becky | b. Always by himself, a loner |
| _____ 3. Jim | c. Popular, outgoing, fun-loving |
| _____ 4. Huck | d. Always critical |
| _____ 5. Mrs. Watson | e. Sneaky, lying, scheming |
| | f. Kind, gentle, loving |
| | g. Dull, slow-moving |

OBJECTIVE 3: The student can write a plausible alternative ending to a story.

TEST ITEM: You have just read the story *Huckleberry Finn*. In 40 words, write a different ending to the story that would be believable.

OBJECTIVE 4: The student will recognize whether certain events occurred.

TEST ITEM: Below is a list of incidents in *Huckleberry Finn*. Circle T if it happened in the story and F if it did not.

- | | | |
|---|---|---|
| 1. The thieves were killed in the storm on the river. | T | F |
| 2. Jim gained his freedom. | T | F |
| 3. Tom broke his leg. | T | F |

Objective 1 was measured using a multiple-choice format. We might also have tested this objective using a true-false or matching format. Similarly, **objective 2** lends itself to a multiple-choice as well as a

matching format. Since many circumstances alternative item formats may be appropriate, the choice between them will be made on the basis of other considerations. For example, time constraints or your preference for, or skill in, writing^ different types of items will undoubtedly influence your choice of item format. However, **objective 3** requires an essay item. There is no way this objective can be measured with an objective item. Similarly, **objective 4** lends itself almost exclusively to a single format. Perhaps other formats would work, but the true-false format certainly does the job. In short, there are times our objectives tell us which format to use. At other times we must consider other factors. Now let's look more closely at the different item formats.

True-False Items

True-false items are popular probably because they are quick and easy to write, or at least they seem to be. Actually, true-false items do take less time to write than good objective items of any other format, but *good* true-false items are not all that easy to write. Consider the following true-false items. Use your common sense to help you determine which items are good and which are poor.

EXERCISE: Put a **G** in the space next to the items you believe are good true-false items and a **P** next to the items you feel are poor.

1. High-IQ children always get high grades in school. ____
2. Will Rogers said, "I never met a man I didn't like." ____
3. If a plane crashed on the Mexican-American border, half the survivors would be buried in Mexico and half in the United States. ____
4. The use of double negatives is not an altogether undesirable characteristic of diplomats and academicians. ____
5. Prayer should not be outlawed in the schools. ____
6. Of the objective items, true-false items are the least time-consuming to construct. ____
7. The trend toward competency testing of high-school graduates began in the late 1970s and represents a big step forward for slow learners. ____

Answers: 1. *P*, 2. *G*, 3. *P*. 4. *P*; 5. *P*; 6. *G*; 7. *P*.

In **item 1**, the word *always* is an absolute term. To some extent, true-false items depend on absolute judgments. However, statements or facts are seldom *completely true* or *completely false*. Thus, an alert student will usually answer "false" to items that include *always*, *all*, *never*, or *only*.

To avoid this problem, avoid using terms like *all*, *always*, *never*, or *only*. Item 1 could be improved by replacing *always* with a less absolute term, perhaps *tend*. Thus, item 1 might read:

High-IQ children tend to get high grades in school.

Item 2 is a good one. To answer the item correctly, the students would have to know whether Will Rogers made the statement. Or do they? Consider the following situation:

Mrs. Allen, a history teacher and crusader against grade inflation, couldn't wait to spring her latest creation on her students. She had spent weeks inserting trick words, phrases, and complicated grammatical constructions into her 100-item true-false test. In order to ensure low grades on the test, she allowed only 30 minutes for the test. Although her harried students worked as quickly as they could, no one completed more than half the items, and no one answered more than 40 items correctly. No one, that is, except Tina. When Tina handed her test back to Mrs. Allen after 2 minutes, Mrs. Allen announced, "Class, Tina has handed in her test! Obviously, she hasn't read the questions and will earn a zero!" However, when she scored the test, Mrs. Allen was shocked to see that in fact Tina had answered 50 items correctly. She earned the highest score on the test without even reading Mrs. Allen's tricky questions. Confused and embarrassed, Mrs. Allen told the class they would have no more true-false tests and would have essay tests in the future.

This points to the most serious shortcoming of true-false items: With every true-false item, regardless of how well or poorly written, the student has a 50 percent chance of guessing correctly even without reading the item! In other words, on a 50-item true-false test, we would expect individuals who were totally unfamiliar with the content being tested to answer about 25 items correctly. However, this doesn't mean you should avoid true-false items entirely, since they are appropriate at times.

1. Encourage *all* students to guess when they do not know the correct answer. Since it is virtually impossible to prevent certain students from guessing. Encouraging all students to guess should equalize the effects of guessing. The test scores will then reflect a more or less equal "guessing factor" *plus* the actual level of each student's knowledge. This will also prevent test-wise students from having an unfair advantage over non-test-wise students.
2. Require revision of statements that are false. With this approach, space is provided for students to alter false items to make them

true. Usually the student also underlines or circles the false part of the item. Item 1 is revised below along with other examples.

T (F) High-IQ children always get high grades in school. (*tend to*)

T (F) Panama is north of Cuba. (*south*)

T (F) September has an extra day during leap year. (*February*)

With such a strategy full credit is awarded only if the revision is correct. The disadvantage of such an approach is that more test time is required for the same number of items and scoring time is increased.

Item 3 is a poor item, but Mrs. Allen would probably like it because it is a trick question. "Survivors" of a plane crash are *not* buried! Chances are that you never even noticed the word *survivors* and probably assumed the item referred to fatalities. Trick items may have a place in tests of critical reading or Visual discrimination (in which case they would no longer be trick questions), but seldom are they appropriate in the average classroom test. Rewritten, item 3 might read:

If a plane crashes on the Mexican-American border, half the fatalities would be buried in Mexico and half in the United States

Item 4 is also poor. First of all, it includes a double negative—*not* and *undesirable*. Items with a single negative are confusing enough. Negating the first negative with a second wastes space and test-taking time and also confuses most students. If you want to say something, say it positively. The following revision makes this item slightly more palatable.

The use of double negatives is an altogether desirable trait of diplomats and academicians.

We said slightly more palatable because the item is still troublesome. The word *altogether* is an absolute, and we now know we should avoid absolute since there usually are exceptions to the rules they imply. Thus, when we eliminate *altogether* the item reads:

The use of double negatives is a desirable trait of diplomats and academicians.

However, the item is still flawed because it states an opinion, not a fact. Is the item true or false? The answer depends on whom you ask. To most of us, the use of double negatives is probably undesirable, for the reasons already stated. To some diplomats, the use of double negatives may seem highly desirable. In short true-false statements should normally be used to measure knowledge of fact

information. If you must use a true-false item to measure knowledge of an opinionated position or statement, state the referent (the person or group that made the statement or took the position), as illustrated in the following revision:

According to the National Institute of Diplomacy, the use of double negatives is a desirable trait of diplomats and academicians.

Item 5 further illustrates this point. It is deficient because it states an opinion. It is neither obviously true nor obviously false. The revision below includes a referent which makes it acceptable.

The American Civil Liberties Union (ACLU) has taken the position that prayer should *not be* outlawed in the schools.

Notice the word *not* in item 5. When you include a negative in a test item, highlight it in italics, underlining, or uppercase letters so the reader will not overlook it. Remember that, unlike Mrs. Allen, you intend to determine whether your students have mastered your objective, not to ensure low test scores.

Item 6 represents a good item. It measures factual information, and the phrase "Of the objective items" qualifies the item and limits it to a specific frame of reference.

The last **item** is deficient because it is double-barreled. It is actually two items in one. When do you mark *true* for a double-barreled item? When both parts of the item are true? When one part is true? Or only when the most important part is true? The point is that items should measure a single idea. Double-barreled items take too much time to read and comprehend. To avoid this problem, simply construct two items, as we have done here:

The trend toward competency testing of high-school graduates began in the late 1970s. The trend toward competency testing represents a big step forward for slow learners.

Better? Yes. Acceptable? Not quite. The second item is opinionated. According to whom is this statement true or false? Let's include a referent.

According to the Office of Education, the trend toward competency testing of high-school graduates is a big step forward for slow learners.

Whose position is being represented is now clear, and the item is straightforward.

Suggestions for Writing True-False Items

1. The desired method of marking true *or false* should be clearly explained before students begin the test.
2. Construct items that are definitely true or definitely false, without additional qualifications. If opinion is used, attribute it to some source.
3. Use relatively short statements and eliminate extraneous material.
4. Keep true and false statements at approximately the same length, and be sure that there are approximately equal numbers of true and false items
5. Avoid using double-negative statements. They take extra time to decipher and are difficult to interpret.
6. Avoid the following:
 - a. verbal clues, absolutes, and complex sentences
 - b. broad general statements that are usually not true or false without further qualifications
 - c. terms denoting indefinite degree (for example, *large, long time, regularly*), or absolutes (for example, *never, only, always*)
 - d. placing items in a systematic order (for example. TTFF, TFTF, and so on)
 - e. taking statements directly from the text and presenting them out of context

Matching Items

Like true-false items, matching items represent a popular and convenient testing format. Just like good true-false items, though, good matching items are not as easy to write as you might think. Imagine you are back in your tenth-grade American History class and the following matching item shows up on your test. Is it a good matching exercise or not? If not, what is wrong with it?

| <i>A</i> | <i>B</i> |
|---------------|---|
| 1. Lincoln | a. President during the twentieth century |
| 2. Nixon | b. Invented the telephone |
| 3. Whitney | c. Delivered the Emancipation |
| Proclamation | |
| 4. Ford | d. Recent president to resign |
| from office | |
| 5. Bell | e. Civil-rights leader |
| 6. King | f. Invented the cotton gin |
| 7. Washington | g. Our first president |

8. Roosevelt

more than two terms

h. Only president elected for

See any problems? Compare those you have identified with the list of faults and explanations below.

Homogeneity.

The lists are *not* homogeneous. Column A contains names of presidents, inventors and a civil-rights leader. Unless specifically taught as a set of related public figures or ideas, this example represents too wide a variety for a matching exercise. To prevent this from happening you might title your lists (for example, "United States Presidents"). This will help keep irrelevant or filter items from creeping in. If you really want to measure student knowledge of presidents, inventors, and civil-rights leaders, then build three separate matching exercises. Doing so will prevent implausible options from being eliminated by the student. When students can eliminate implausible options they are more likely to guess correctly. For example, the student may not know a president who resigned from office, but may know that Washington and Lincoln were presidents, and that neither was recent. Thus, the student could eliminate two options, increasing the chance of guessing correctly from once out of eight to one out of six.

Order of Lists.

The lists should be reversed; that is, column A should be column B, and column B should be column A. This is a consideration to save time for the test taker. We are trained to read from left to right. When the longer description is in the left-hand column the student only reads the description once and glances down the list of names to find the answer. As the exercise is now written, the student reads a name and then has to read through all or many of the more lengthy descriptions to find the answer, a much more time-consuming process.

Easy Guessing.

There are equal numbers of options and descriptions in each column. Again, this increases the chances of guessing correctly through elimination. In the preceding exercise, if a student did not know who invented the cotton gin but knew which of the names went with the other seven descriptions, the student would arrive at the correct answer through elimination. If there are at least three more options than descriptions, the chances of guessing correctly in such a situation are reduced to one chance in four. Alternatively, the instructions for the exercise may be written to indicate that each option *may* be used more than once.

Poor Directions.

Speaking of directions, those included above are much too brief. Matching directions should specify the basis for matching. For example:

DIRECTIONS: Column A contains brief descriptions of historical events. Column B contains the names of presidents. Indicate which man was president when the historical event took place by placing the appropriate letter to the left of the number in column A.

The directions also do not indicate *how* the matches should be shown. Should lines be drawn? Should letters be written next to numbers, or numbers next to letters? Failure to indicate how matches should be marked can greatly increase your scoring time.

Too Many Correct Answers

The description "President during the twentieth century" has three defensible answers: Nixon, Ford, and Roosevelt. You say you meant Henry Ford, inventor of the Model T, not Gerald Ford! Well, that brings us to our final criticism of this matching exercise.

Ambiguous Lists.

The list of names is ambiguous. Franklin Roosevelt or Teddy Roosevelt? Henry Ford or Gerald Ford?

When using names, always include first and last names to avoid such ambiguities.

Now that we have completed our analysis of this test item, we can easily conclude that it needs revision. Let's revise it, starting by breaking the exercise into homogeneous groupings.

DIRECTIONS: Column A describes events associated with United States presidents. Indicate which name in Column B matches each event by placing the appropriate letter to the left of the number of column A. Each name may be used only once.

| Column A | Column B |
|---|-----------------|
| --- 1. Only president not elected to office | |
| a. Abraham Lincoln | |
| --- 2. Delivered the emancipation proclamation | b. |
| Richard Nixon | |
| --- 3. Only president to resign from office | c. |
| Gerald Ford | |
| --- 4. Only president elected for more than two terms | d. |
| George Washington | |
| --- 5. The first president of the United States | e. |
| Franklin Roosevelt | |
| f. Theodore Roosevelt | |

g. Thomas Jefferson

h. Woodrow Wilson

We can make one more clarification. It is a good idea to introduce some sort of order—chronological, numerical, or alphabetical—to your list of options. This saves the reader time. Students usually go through the list several times in answering a matching exercise, and it is easier to remember a name's or date's location in a list if it is in some sort of order. Thus, we can arrange the list of names in alphabetical order to look like this:

| Column A | Column B |
|--|----------|
| --- 1. Only president not elected to office a. Gerald Ford | |
| --- 2. Delivered the emancipation proclamation Thomas Jefferson | b. |
| --- 3. Only president to resign from office Abraham Lincoln | c. |
| --- 4. Only president elected for more than two terms Richard Nixon | d. |
| --- 5. The first president of the United States Franklin Roosevelt | e. |

f. Theodore Roosevelt

g. George Washington

h. Woodrow Wilson

Our original exercise contained two items relating to invention. If we were determined to measure only knowledge of inventors through a matching exercise, we would want to add at least one more item. Normally, at least three items are used for matching exercises. Such an exercise might look like the following:

| Column A | Column B |
|--|-------------------|
| --- 1. Invented the cotton gin Graham Bell | a. Alexander |
| --- 2. One of his inventions was the telephone | b. Henry Bessemer |
| --- 3. Famous for inventing the wireless | c. Thomas Edison |
| | d. |
| Guglielmo Marconi | |

e. Eli

Whitney

f. Orville

Wright

Notice we have complete directions, there are three more options than descriptions, the lists are homogeneous, and the list of names is alphabetically ordered. But what about the final item remaining from our original exercise? Let's say we want to determine whether our students know that Martin Luther King, Jr., was a civil-rights leader. We can construct another matching exercise with one column listing the names of civil-rights leaders and another listing civil-rights accomplishments. However, an alternative would be simply to switch item formats. Usually, single items that are removed from matching exercises because of non-homogeneity are easily converted into true-false, completion, or, with a little more difficulty, multiple-choice items. For example:

True-False T F Martin Luther King, Jr.,
was a civil-rights leader.

Completion The name of the black civil-rights leader
assassinated in 1968 was ____ . *Multiple-Choice*

Which of the following was a civil-rights leader?

- a. Jefferson Davis
- b. Martin Luther King, Jr.
- c. John Quincy Adams
- d. John Wilkes Booth

Suggestions for Writing Matching Items

1. Keep both the list of descriptions and the list of options fairly short and homogeneous—they should both fit on the same page. Title the lists to ensure homogeneity, and arrange the descriptions and options in some logical order. If this is impossible you're probably including too wide a variety in the exercise. Try two or more exercises.

2. Make sure that all the options are plausible distractors for *each* description to ensure homogeneity of lists.

3. The list of descriptions should contain the longer phrases or statements, while the options should consist of short phrases, words, or symbols.

4. Each description in the list should be numbered (each is an item), and the list of options should be identified by letter.

5. Include more options than descriptions. If the option list is longer than the description list, it is harder for students to eliminate

options. If the option list is shorter, some options must be used more than once. Always include some options that do not match any of the descriptions, or some that match more than one, or both.

6. In the directions, specify the basis for matching and whether options can be used more than once.

Multiple-Choice Items

Another popular item format is the multiple-choice question. Practically every-one has taken multiple-choice tests at one time or another, but probably more often in high school and college than elementary school. This doesn't mean that multiple-choice items are not appropriate in the elementary years; it suggests only that one needs to be cautious about using them with younger children.

As before, let's start by using common sense to identify good and poor multiple-choice items in the following exercise:

EXERCISE; Place a **G** next to a good item and a **P** next to a poor item.

- 1. U.S. Grant was an
 - a. president.
 - b. man.
 - c. alcoholic.
 - d. general.
- 2. In what year did humans first set foot on the moon?
 - a- 1975
 - b. 1957
 - c. 1969
 - d. 1963
- 3. Free-floating structures within the cell that synthesize protein *are* called
 - a. chromosomes.
 - b. lysosomes.
 - c. mitochondria.
 - d. free ribosomes.
- 4. The principal value of a balanced diet is that it
 - a. increases your intelligence.
 - b. gives you something to talk about with friends.
 - c. promotes mental health.
 - d. promotes physical health
 - e. improves self-discipline.

Go over the exercise again. Chances are you'll find a few more problems the second time. Here's the answer key and a breakdown of the faults found in each item.

Answers: 1. P; 2. G; 3. P; 4. P.

Most students would probably pick up on the "grammatical clue" in the first item. The article "an" eliminates options a, b, and d immediately, since "U.S. Grant was an man," "an president," or "an general" are not grammatically correct statements. Thus, option c is the only option that forms a grammatically correct sentence. Inadvertently providing students with grammatical clues to the correct answer is very common in multiple-choice items. The result is decreased test validity. Students can answer items correctly because of knowledge of grammar, not content.

Replacing "an" with "a/an" would be one way to eliminate grammatical clues in your own writing. Other examples would be "is/are," "was/were," "h her," and so on. As an alternative, the article, verb, or pronoun may be included the list of options, as the following example illustrates:

Poor: Christopher Columbus came to America
in a

- a. car.
- b. boat.
- c. airplane.
- d. balloon.

Better: Christopher Columbus came to
America in

- a. a car.
- b. a boat.
- c. an airplane.
- d. a balloon.

Let's return to the first item again and replace "an" with "a/an":

U. S. Grant was a/an

- a. president.
- b. man.
- c. alcoholic.
- d. general.

There! We've removed the grammatical clue, and we now have an acceptable item, right? Not quite. We now have an item free of grammatical clues, but still seriously deficient. What is the correct answer?

This item still has a serious flaw: multiple defensible answers. In fact four options are defensible answers! U. S. Grant was a president, a man, a general, as historians tell us, an alcoholic. Including such an item on a test contribute nothing to your understanding of student knowledge. But what you do when you have

an item with more than one defensible answer? The answer of course, is to eliminate the incorrect but defensible option or options.

Let's assume **item 1** was written to measure the following objective:

The student will discriminate among the United States presidents immediately before, during, and after the United States Civil War.

We could modify item 1 to look like this:

- U. S. Grant was a
- a. general.
 - b. slave.
 - c. pirate.
 - d. trader

This item is fine, from a technical standpoint. The grammatical error has been eliminated and there is but one defensible answer. However, it does *not* match the instructional objective; it is not very valuable as a measure of student achievement of the objective. We could also modify the item to look like this:

Of the following, who was elected president after the Civil War?

- a. U. S. Grant
- b. Andrew Johnson
- c. Abraham Lincoln
- d. Andrew Jackson

This item is technically sound, and all response alternatives are relevant to the instructional objective. It meets the two main criteria for inclusion in a test:

The item is technically well constructed, and it matches the instructional objectives.

We said **item 2** was good, but it can still stand some improvement. Remember when we recommended arranging lists for matching items in alphabetical or chronological order? The same holds true for multiple-choice items. To make a good item even better, arrange the options in chronological order. Revised, the item should look like this:

In what year did humans first set foot on the moon?

- a. 1957
- b. 1963
- c. 1969
- d. 1975

The major deficiency in item 3 is referred to as a "stem clue." The statement portion of a multiple-choice item is called the *stem*, and the correct answer and incorrect choices are called *options* or *response alternatives*. A stem clue occurs when the same word or a close

derivative occurs in both the stem and options, thereby clueing the test taker to the correct answer. **In item 3** the word *free* in the option is identical *to free* in the stem. Thus, the wise test taker has a good chance of answering the item correctly without mastery of the content being measured. This fault can be eliminated by simply rewording the item without the word *free*.

The structures within the cell that synthesize protein are called

- a. chromosomes.
- b. lysosomes.
- c. mitochondria.
- d. ribosomes.

Item 4 is related to the "opinionated" items we considered when we discussed true-false items. Depending on the source, or referent, different answers may *be* the "right" answer. To Person X, the principal value may be to promote physical health; to Person Y, the principal value may be to improve self-discipline. As stated earlier, when you are measuring a viewpoint or opinion, be sure to state the referent or source. To be acceptable the item should be rewritten to include the name of an authority:

The USDA feels the principal value of a balanced diet is that it

- a. increases your intelligence.
- b. gives you something to talk about.
- c. promotes mental health.
- d. promotes physical health.
- c. improves self-discipline.

General Guidelines for Writing Test

Items

- Begin writing items far enough in advance that you will have time to revise them.
- Match items to the proper difficulty level to ensure valid measure of instructional objectives.
- Be sure each item deals with an important aspect of the content area and not with trivia (secondary or less important aspects)
- Be sure that the problem posed is clear and unambiguous.
- Be sure that each item is independent of all other items. The answer to one item should not required as a condition for answering the next item.
- Be sure the item has one correct or best answer on which experts would agree.

- Avoid replication from textbook...don't quote directly from textual materials.
- Avoid trick or catch-questions your tests.

Conclusion

It is now clear that designing a test is not the simple task that one thinks it to be. As a matter of fact, the choice of the item format is sometimes determined by our instructional objective. At other times other factor should influence our choice, advantages and disadvantages of the formats in question notably. Objectivity in testing is therefore approached when teachers are able to construct well-designed tests which contribute in yielding valid and reliable measures of how well they teach and how well their students learn.

References:

- Douglas, K. (2004) *Teacher ideas on teaching and testing English language learners: Summary of focus group discussions (C-SAVE Rep. No. 223)*. Maddison, WI: University of Wisconsin, Center for the Study of Assessment Validity and Evaluation.
- Elliott, S.N., Kratochwill, T.R., and Gilbertson-Schulte, A. (1999) *Assessment accommodations guide*. Monteray, CA: CTB McGraw-Hill.
- Kubiszyn, T. and Borich, G. (1993) *Educational Testing and Measurement*. Harpercollins College Publishers. Texas.
- Shepard, L.A. (1990) Inflated test score gains: Is the problem old norms of teaching the test? *Educational Measurement: Issues and Practice*, 9(3), 15–22.