

ISSN 1560—2907

Principles for Writing Multiple-Choice Items  
in Vocabulary, Grammar, and Reading Tests:  
A Study on the Test Making Process of an  
Achievement Test at Fooyin Institute of Technology

Mei-er Yu Yi-ching Lin

宜蘭技術學報 第八期

中華民國九十一年六月

# Principles for Writing Multiple-Choice Items in Vocabulary, Grammar, and Reading Tests: A Study on the Test Making Process of an Achievement Test at Fooyin Institute of Technology

Mei-er Yu, Yi-ching Lin

Lecturers, Department of Foreign Languages, Fooyin Institute of Technology

## Abstract

The purpose of this paper is to outline certain principles for EFL teachers to follow in constructing multiple-choice items in vocabulary, grammar, and reading tests. Item examples are derived from the teacher-made achievement test which is intended to be administered at Fooyin Institute of Technology in the spring semester, 2001. We start by reviewing types of language tests and multiple choice. We then provide guidelines for constructing multiple-choice items in vocabulary, grammar, and reading tests and introduce the upcoming achievement test at Fooyin. We further present the results that the teachers still construct poor items even though the guidelines have been provided in advance. Finally, we propose the limitations of multiple-choice items and suggest the need to offer EFL teachers in-service training programs about language testing.

**Key Words** : achievement test, distractor, option

## 單字、文法、閱讀測驗選擇題命題常犯錯誤之探討:以輔英技術學院教師自製成就測驗為例

余美娥、林宜靜

輔英技術學院應用外語科講師

## 摘 要

語言測驗在「教」與「學」上扮演舉足輕重的角色，它不但可評估學生的學習成果，也提供了教師檢討教學內容及方法的依據。本文目的在介紹語言測驗之種類與選擇題之特性，並進一步探討單字、文法、閱讀測驗選擇題命題之原則，引用的例子皆為作者在輔英技術學院預計於 89 學年度下學期實施的教師自製成就測驗中所得之試題，文中也從所得之試題中分析命題常犯之錯誤。最後，本文並提供了英文教師使用選擇題時應注意之要項並給予建議。

**關鍵詞**：成就測驗、誘答選項、選擇項

## I. Introduction

Language testing is closely interrelated with teaching and learning. Well-constructed tests can assess students' abilities with regard to the course objectives and can diagnose their weaknesses and strengths. They also provide teachers the information to evaluate the effectiveness of teaching and can serve as a basis for instruction improvement. On the other hand, tests with poor quality can not measure students' abilities objectively and may have harmful backwash on teaching and learning. As a result, teachers should be cautious with writing test items.

An achievement test which consists of 50 multiple-choice items is intended to be administered at Fooyin Institute of Technology for the five-year college students majoring in nursing and medical technology in the spring semester, 2001. A committee set up by several EFL teachers will implement the teacher-made achievement test. The committee provides the EFL teachers the principles for constructing multiple-choice items and evaluates the test items collected from them. However, as members of the committee, we still observe some common errors in writing multiple-choice items. It appears that the teachers still construct poor items even though the guidelines have been supplied in advance and that the common errors actually occur if teachers do not follow these principles carefully. To avoid impropriety, teachers should be discreet in writing test items.

## II. Types of Language Tests

Language tests can be classified into four categories by purpose: placement, diagnostic, achievement, and proficiency. A placement test, as the name suggests, is designed to assign students to the particular level of a teaching program most appropriate to their abilities. A diagnostic test is intended to pinpoint students' strengths and weaknesses in learning progress. Teachers should adjust their teaching and determine what further instruction is necessary. Diagnostic tests are also useful for individualized instruction or self-instruction to help students perceive where to focus their efforts (Hughes, 1989).

An achievement test is directly based on course objectives and is typically executed at the end of a course to assess how effectively students have mastered the instructional objectives. Teachers can learn students'

achievements and the adequacy of the course from achievement tests (Brown, 1996). Compared with diagnostic tests, achievement tests cover a much broader range of materials and are concerned with long-term rather than short-term objectives (Harrison, 1983). Finally, the aim of a proficiency test is to measure the student's ability or proficiency in a language. Therefore, the test is not based on the content or objectives of courses that he or she may have previously taken.

The EFL teachers at Fooyin intend to know how successfully students have reached the course objectives and evaluate the appropriateness of the language program. Hence, the test which is planned to be implemented at Fooyin is considered as an achievement test.

## III. Multiple-Choice Items

A multiple-choice item consists of a *stem*, which is stated as a direct question or an incomplete statement, and a number of suggested solutions called *options* (also called *choices* or *alternatives*). The correct option is *answer* and the remaining ones are the *distractors*, which are incorrect but plausible. The purpose of the distractors is to distract those students who are unsure or in doubt about the correct answer. There are either three, four, or five options but four options are preferred for the reason that two distractors will increase the chances of guessing the correct answer and a fourth distractor is difficult to devise.

Among item types, multiple choice has been recommended and used for the testing of many language abilities (Hughes, 1989). It is widely used by teachers, schools, and assessment organizations and is one of the most commonly applicable test techniques to assess achievement (Linn & Gronlund, 1995). The chief advantage of multiple choice tests is that scoring can be simple, rapid, and economical. They can be scored objectively and thus make the test more reliable than subjectively scored tests (Bailey, 1998). Although multiple-choice items test only recognition knowledge rather than the ability to use language, they are confirmed effective in assessing student's ability to recognize correct grammatical forms, vocabulary, etc. and to make important discrimination in the target language. Consequently, both the student and the teacher can identify areas of difficulty by analyzing student's performance on the multiple-choice test (Heaton, 1988). Another advantage of multiple-choice items is that they can undergo pre-tests fairly easily. It is usually possible to assess in advance the

difficulty level of each item and that of the test as a whole (Weir, 1990).

Nevertheless, a successful multiple-choice item is difficult and time-consuming to construct. Since poorly made tests can not measure students' abilities objectively and may have harmful backwash on teaching and learning, it is essential to follow certain principles in constructing multiple-choice items.

#### IV. Principles for Constructing Multiple-Choice Items

The principles below are based on the guidelines presented by Madsen (1983), Harris (1969), Henning (1982 & 1987), and Linn & Gronlund (1995).

##### Vocabulary

One of the vocabulary test types in the upcoming achievement test at Fooyin is multiple-choice completion, in which a word is deleted from a sentence and the candidate selects a correct answer depending on context clues and sentence meaning. Another vocabulary test type in the achievement test is multiple-choice paraphrase, in which a lexical item is underlined in context and testees are required to choose the best synonym or paraphrase of the lexical item depending more on knowing the lexical item than on detecting meaning in the sentence context. Following are the guidelines for writing vocabulary items and the examples are taken from the test items gathered from the EFL teachers to compose the achievement test:

1. Distractors should be the same form of word as the correct answer (Madsen, 1983).

Ex: The \_\_\_\_\_ Thanksgiving dinner consists of turkey, cranberry sauce, several vegetables, and pumpkin pie.

A. opinion B. instead C. typical D. prepare

Students may perceive an adjective is necessitated in the item and *typical* is the only adjective among the options; thus, students might answer the item correctly without knowing the meaning of the answer.

2. Grammatical clues should be avoided in the stem (Madsen, 1983).

Ex: My boyfriend gave me an \_\_\_\_\_ watch. It was made 100 years ago.

A. cruel B. antique C. sensitive D. liquid

Examinees who know the article system in English can recognize *antique* is correct because *an* must be

followed by a word beginning with a vowel sound. The item could be corrected by either adding *a* in the stem:

My boyfriend gave me a(an) \_\_\_\_\_ watch. It was made 100 years ago.

A. cruel B. antique C. sensitive D. liquid

or removing the article from the stem to the options:

My boyfriend gave me \_\_\_\_\_ watch. It was made 100 years ago.

A. a cruel B. an antique C. a sensitive D. a liquid

3. All the options should be on nearly the same level of difficulty (Harris, 1969).

Ex: You can't touch the paper because your hands are \_\_\_\_\_. It should be dry and neat.

A. clean B. hot C. cold D. moist

Obviously distractors A, B, and C of this item could be eliminated because they are much easier than the correct answer, *moist*, and not fit in the context. Consequently, candidates would select the right answer simply by the process of elimination without knowing the meaning of the correct option.

4. All the options should avoid needless redundancy (Henning, 1987).

Ex: I dress in a conservative way.

A. My clothes are fabulous in style.  
B. My clothes are fashionable in style.  
C. My clothes are appropriate in style.  
D. My clothes are conventional in style.

Testees are required to read the repetition of redundant material throughout the alternatives. Therefore the test would be inefficient in that testees need much time to derive the information available from a given period of time for testing (Henning, 1982). A better format for such an item would be:

I dress in a conservative way.

A. fabulous B. fashionable  
C. appropriate D. conventional

5. Avoid using distractors with similar meanings (Madsen, 1983).

Ex: The party guests were very amazed by Christina's behavior.

A. pleased B. surprised C. cheerful D. bored

Students who recognize *pleased* and *cheerful* are close in meaning might eliminate both and discern either *surprised* or *bored* can be the correct answer.

6. Avoid using a pair of words with opposite meanings as distractors (Madsen, 1983).

Ex: When we got to John's house, we found it was as silent as the grave.

- A. quiet B. loud C. clean D. dark

If students perceive *quiet* and *loud* as antonyms, they might eliminate the last two options and select one of the antonyms as the right answer.

7. Each item should contain only one correct or obviously best answer (Linn & Gronlund, 1995).

Ex: Even though I failed in the test, my English teacher still \_\_\_\_\_ me to work harder.

- A. respected B. employed  
B. C. encouraged D. persuaded

Both *encouraged* and *persuaded* can be correct answers. This error can be avoided by having another EFL teacher read through all the items before using them on a test.

8. All the options should be nearly the same length (Harris, 1969).

Ex: They wanted to ignore the children.

- A. pay no attention to B. improve  
C. take away D. please

There is a tendency that the longest option is the obvious answer because it is often necessary to paraphrase the lexicon item sufficiently. Care must be taken that the correct answer should not be generally the longest.

### Grammar

The type of grammar test used in the achievement test is multiple-choice completion, which presents an incomplete sentence stem followed by four multiple-choice alternatives for completing the sentence. Principles for writing grammar items are as follows.

1. Avoid using distractors which are pronounced alike (Madsen, 1983).

Ex: Tony \_\_\_\_\_ go swimming but not any more now.

- A. used to B. use to C. was used to D. am used to

Both A and B would sound alike when normally spoken. The choice of the correct answer turns to be a spelling problem and this "error" would appear in writing but not in speech. It might be used on a writing test, but not appropriate on a grammar test (Harris, 1969).

2. Avoid using mixed options (Henning, 1987).

The following item is used to test tense but the options

confuse the notion of tense with that of agreement.

Ex: My first year in the United States, I met a girl who \_\_\_\_\_ my classmate in Taiwan.

- A. have been B. being C. had been D. were

Examinees might disregard A and D due to lack of subject-verb agreement, not on the basis of impropriety of tense.

3. Avoid nonsense distractors (Henning, 1987).

Ex: My parents were being old-fashioned and \_\_\_\_\_ as Americans should behave.

- A. not behaving B. behaving not  
C. did not behaved D. not behaved

C contains structure that does not occur in grammatically acceptable English. Nonsense distractors tend to be weak distractors and students may learn errors from the test itself. Such distractors might have negative washback on instruction and they should be avoided for this reason (Henning, 1987).

4. Avoid using inconsistent distractors (Henning, 1987).

Ex: I consider \_\_\_\_\_ rude to speak loudly in the hotel lobby.

- A. this B. that C. it D. its

A, B, and C are all pronouns in this item. Test-wise students may recognize D is the only possessive adjective and might eliminate this option at the start. Consequently, the item has only three effectual options which make it a simpler item than originally intended.

### Long Passage Reading Comprehension

Readings are generally built around the grammatical structures and lexical items (Harris, 1969) and passage reading comprehension is the most integrative and challenging type of reading test (Madsen, 1983). The proper length of longer passage is between 100 to 300 words and the passages used in the achievement test will run from 150 to 200 words. All the passages accompanied with four to six questions are adopted by EFL teachers from external sources though, there are still several principles to follow for either selecting or writing appropriate items.

1. Selection of the correct answer should not merely match the words in the options with the same words in the passage (Harris, 1969).

An example would be the following:

*For the temporary relief of minor aches and pains, take one capsule every six hours while symptoms persist. If pain does not*

*respond to one capsule, two capsules may be used but do not take more than six capsules in twenty-four hours, unless told by a doctor. The smallest effective dose should be used. Take with food or milk if occasional upsets stomach or stomach pain occurs with use.*

Ex: If the product upsets your stomach, you may

- A. take a rest
- B. take it with milk or food
- C. take it with aspirin
- D. take one capsule every six hours

Students can select the correct answer by purely matching of words: *upset stomach* and *take with food or milk* are the phrases used in both the passage and the test item. The correct answer should be paraphrased to avoid superficial testing of comprehension (Harris, 1969).

2. Items should require comprehension of the passage. It should not be feasible to answer items correctly purely on the basis of general knowledge (Henning, 1982).

Ex: The polar bear lives \_\_\_\_ .

- A. at the South Pole
- B. in warm weather
- C. near the North Pole
- D. on land

The answer is general knowledge for most people regardless of the content of the passage. Examinees can therefore choose the answer correctly without comprehending the passage.

3. All the options should be roughly the same length.

Ex: How did she know this young man is going to the library?

- A. The young man told her.
- B. The young man told someone else and she accidentally heard of it.
- C. The roommate asked him about it.
- D. The author told her roommate about it.

As stated in the principles for vocabulary item, the longest and most lucid option is frequently the correct answer. Candidates who neither read nor comprehend the passage might choose B merely because greater detail is provided.

## V. Methodology

To measure how successfully students have achieved the instructional objectives after the three-year language course and evaluate the effectiveness of the language

program, the EFL teachers at Fooyin Institute of Technology intend to administer an achievement test for the five-year college students majoring in nursing and medical technology in the spring semester, 2001. In view of the wide content and not to burden the non-English major students, the test will consist of two parts as a midterm and final examination. Part I will be based on topics covered in the first year, and Part II will include the second and the third years' instruction. The textbooks for the first year course are *Our Own Stories* and *Impact Intro. News for Now* and *First Impact* are textbooks for the second-year students and *Focus on Topics* is for the third graders.

A committee is set up to implement the teacher-made achievement test and evaluate the test items collected from EFL teachers instructing target students. The test items turned in by the EFL teachers must follow the principles furnished by the committee to ensure validity and reliability of the tests. The committee will also select items from the testing pool to compose the achievement test.

Eight sections are included in both parts of the test: vocabulary in context, grammar, word forms, dialogues, single sentence comprehension, cloze, short passage comprehension, and long passage reading comprehension. The two parts of the test will consist of 50 multiple-choice items respectively; one correct answer among four options. Only vocabulary, grammar, and long passage reading comprehension will be discussed in this paper.

## VI. Results and Discussion

The test items obtained from the EFL teachers are analyzed and the results are shown in the following tables.

Table 1. Distribution of collected test items

Sections	Total number of test items	Number of improper test items	Percentage of improper test items
Vocabulary	456	194	42.5%
Grammar	350	37	10.6%
Reading comprehension	205 (45 passages)	52	25.4%

The result indicates that in constructing test items the teachers made much more mistakes in vocabulary section than in grammar section. As for the reading comprehension, although all the passages and the test items are adopted from external sources, 25.4% of items are not acceptable.

As shown in Table 2, the violation of principle 3 is much higher than other violations. This violation also

appeared on the 1998 Joint College Entrance Examination in Taiwan (Hsu & Lu, 1998).

In table 3, it is found that 54.1% of items use nonsense or ungrammatical distractors among the improper test items. This error also occurs frequently on classroom tests in junior high schools in Taiwan (Shih, 1985). It also reveals that 21.6% use inconsistent distractors which make the test items easier than originally intended.

Table 2. Distribution of improper items in vocabulary section

Principles	Number of improper items	Percentage among improper items
1.All options should be the same word class	36	18.6%
2.Grammatical clues should be avoided in the stem	19	9.8%
3.All options should be on nearly the same level of difficulty	79	40.7%
4.All options should avoid needless redundancy	4	2.1%
5.Avoid using distractors with similar meanings	13	6.7%
6.Avoid using distractors with opposite meanings	20	10.3%
7.Each item should contain only one correct or obviously best answer	18	9.3%
8.All options should be nearly the same length	5	2.6%
Total	194	100%

Table 3. Distribution of improper items in grammar section

Principles	Number of improper items	Percentage among improper items
1.Avoid using distractors which are pronounced alike	6	16.2%
2.Avoid using mixed options	3	8.1%
3.Avoid nonsense distractors	20	54.1%
4.Avoid using inconsistent distractors	8	21.6%
Total	37	100%

Table 4. Distribution of improper items in reading comprehension section

Principles	Number of improper items	Percentage among improper items
1.Items should not be answered by purely matching of words	29	55.8%
2.Items should not be answered from general knowledge	7	13.5%
3.All options should be roughly the same length	16	30.8%
Total	52	100%

Table 4 indicates that in reading comprehension section, correct answers in 55.8% of improper items are not paraphrased. 13.5% of improper items can be answered from general knowledge and this violation also appeared on the 1994 Joint College Entrance Examination (Chen, 1995).

The results show that although the EFL teachers are provided with principles, they still construct poor test items. It might be partly due to the fact that they did not receive any or enough training in language testing before, thus it might be difficult for them to follow these guidelines.

Even the Joint College Entrance Examination, which is constructed by trained university professors, occasionally contains improper multiple-choice items. Another possibility might be that they devote themselves to teaching, constructing classroom tests, doing research, etc., therefore they did not spend much time reading through these principles and following them. The other reason might be the teacher's personality. Some teachers might not write test items with a serious manner; they might not care if they violate the principles.

## VII. Limitations and Conclusion

Well-constructed English tests can benefit students by helping them master the language when they study for such tests and when tests are returned and discussed. Properly made tests can also help create positive attitudes toward instruction by giving students a sense of accomplishment and a feeling that the teacher's evaluation of them matches what they had been taught (Madsen, 1983). If EFL teachers utilize the principles for constructing multiple-choice items in this paper, much better items will be yielded more quickly.

In spite of certain advantages of multiple-choice items mentioned earlier, EFL teachers should be aware of the limitations of multiple-choice items when using this test technique. It is easy for examinees to cheat or gain scores by guessing. In addition, an examinee might get a multiple-choice item wrong because of some defect in the question which is not revealed on the answer sheet. We will not know whether an examinee's failure is due to lack of comprehension of the text or lack of comprehension of the question. An examinee might get an item correct by eliminating wrong answers, which is a different skill from being able to select the correct answer in the first place (Weir, 1990).

Another limitation of the multiple-choice item is that it tests only recognition knowledge. The performance on a multiple-choice test may not indicate an examinee's productive skill. The student who can identify the correct option in the item may not be able to produce the correct form when speaking or writing. If we are interested in the productive use of language, the test scores may give incomplete information (Hughes, 1989).

The research results reveal that the EFL teachers made much more mistakes in vocabulary section than in grammar section. In reading comprehension section, a number of

items the teachers selected from external sources do not follow the principles. The findings imply that even though the EFL teachers have been provided with the principles for constructing multiple-choice items in advance, the teachers still construct improper items. Language testing plays an important role in both teaching and learning. Well-constructed tests can enhance learning and motivate students. Hence, it is essential that EFL teachers should acquire the knowledge of testing. We should offer EFL teachers inservice teacher-training programs about language testing to allow EFL teachers to get familiar with the principles and techniques and to apply them while constructing test items. As Tang (1993) suggested, we should be well prepared before constructing test items and should not write items casually. He also recommends that we can make reference to the test items written by other teachers and collect good ones. Before using items on a test, it would be better to have another EFL teacher read over to avoid possible errors. As proposed by Chan (1996), even the item type we use is the simplest or most widely used test type, we should construct items with the most cautious manner.

## REFERENCES

1. Bailey, K.M. (1998), *Learning about Language Assessment*, Heinle & Heinle Publishers, Boston, U.S.A.
2. Brown, J.D. (1996), *Testing in Language Programs*, Prentice Hall Regents, New Jersey, U.S.A.
3. Chen, K.T. 陳坤田 (1995), 「八十三學年度大學暨獨立學院入學考試英文試題分析」, 教育研究資訊雙月刊, 第三卷, 第五期, 第 81-102 頁。
4. Chan, L.H. 詹麗馨 (1996), 「國中英語單項式字彙測驗編製的要領」, 英語教學, 第二十一卷, 第一期, 第 61-70 頁。
5. Harris, D.P. (1969), *Testing English as a Second Language*, McGraw-Hill, Inc. New York, U.S.A.
6. Harrison, A. (1983), *A Language Testing Handbook*. Macmillan Publishers Limited, London and Basingstoke, U.K.
7. Heaton, J.B. (1988), *Writing English Language Tests*, Longman Group Limited, New York, U.S.A.
8. Henning, G. (1982), "Twenty Common Testing Mistakes for EFL Teachers to Avoid", *English Teaching Forum*, Vol.20, No.3, pp.33-37.

9. Henning, G. (1987), *A Guide to Language Testing*, Heinle & Heinle Publishers, Boston, U.S.A.
10. Hsu, Y.H. & Lu, F.S. 許亞君、盧富世 (1998), 「八十七學年度大學聯考英文試題簡評」, *英語教學*, 第二十三卷, 第二期, 第 23-40 頁。
11. Hughes, A. (1989), *Testing for Language Teachers*, Cambridge University Press, New York, U.S.A.
12. Linn, R.L. & Gronlund, N.E. (1995), *Measurement and Assessment in Teaching*, Prentice-Hall, Inc, New Jersey, U.S.A.
13. Madson, H.S. (1983), *Techniques in Testing*, Oxford University Press, New York, U.S.A.
14. Shih, Y.H. 施玉惠 (1985), 「國中英語科測驗評量的原則、技巧及態度」, *英語教學*, 第十卷, 第一期, 第 38-49 頁。
15. Tang, T.C. 湯廷池 (1993), 「大學入學考試英文科試題: 檢討與建議」, *第十屆英語文教學論文集*, 民國八十二年七月, 第 649-676 頁。
16. Weir, C.J. (1990), *Communicative Language Testing*, Prentice Hall International (UK) Ltd, England.

90 年 04 月 11 日投稿

90 年 11 月 13 日接受