

PREPARING QUALIFIED MULTIPLE-CHOICE TEST ITEMS: WHAT SHOULD A LANGUAGE TEACHER REALIZE?

By:
Suwartono
Muhammadiyah University of Purwokerto

ABSTRACT

Testing is inseparable from teaching. Occasionally, a teacher has to write test items by himself. At times, a teacher can adapt existing items for his own context. Since testing can give effect on learning (backwash), qualified items are needed. Even, when adoption of items available is chosen, proper consideration remains significant. Facts show that multiple-choice items play an important role in our EFL assessment. This paper aims to raise teacher's awareness about this increasingly popular type of test items.

Good multiple-choice items facilitate learning. Take, for instance, in searching for meaning of new words, the students can be provided with options that include familiar distracters. By doing so, the teacher helps the students find meanings without being nurtured. It also means that the teacher can avoid translation in his effort to help them arrive at meaning. With good multiple-choice items, the teacher gets the students to practice known linguistic items. Qualified multiple-choice items prevent the candidates from disappointment that is likely to result in motivational problem. Imagine when students are suffering from low achievement due to poor items – it's unfair!

Though seemingly simple, multiple-choice test type involves the highest risk of ill construction. Considerations should be made to prepare proper multiple-choice items. In the first place, the options should be equal in length or no option is extremely different. Of course, it is often hard to achieve. In the second, no more than one option is equally acceptable. In addition, the keys should not make a pattern. And, above all, teachers need to conduct item moderation. The latter is also good for promoting the academic atmosphere. To the writer, all those points are the collected wisdom, which is ready to share in the presentation.

Keywords: *qualified, multiple-choice, test items, assessment*

Introduction

Insights into and competence in testing is often neglected by teachers. In the context of education, testing is inseparable from teaching. To some degree, testing affects learning (backwash) either harmfully or beneficially. Of course, what we expect from a testing is a beneficial backwash effect on teaching and learning. Students' dissatisfaction due to poor testing technique or scoring system, for instance, may result in bad attitude towards learning. Accumulation of learned materials with quite rare or no periodic evaluation may hinder rather than help students learn because they do not know the level of their learning mastery. Advance notice of upcoming test will motivate students to learn. Testing with proper level of difficulty might raise or maintain students' positive attitude towards teacher, teaching and learning. Therefore, a good teacher will never ignore test methodology as an integral part of his teaching profession.

In the meanwhile, the fact has shown that most tests in language education across this country involve multiple-choice items. For many years, this type of test item has taken an important role in language testing in our national final examination. The majority or even the entirety of the test items constitutes this test item type. The world's leading language testing program, named "TOEFL" uses almost exclusively this item type.

In daily English classroom practices, multiple-choice can also be used to help students learn better. For example, it has been a long tradition that questions which require longer responses follow the reading passages (most noticeably essay). It is, of course, monotonous and to many students essay items perhaps seem demanding and can lead to reluctance to learn. Multiple-choice items, however, can be an alternative approach to reading or listening comprehension classroom tasks and/or home exercises. In the first place, since the response is limited, multiple-choice items enable the students to concentrate on meaning. Covering things other than meaning in reading or listening comprehension assessment seems unfair! And this will happen if students have to give written or oral responses (productive).

In the second, it is likely that the students will be encouraged to read the texts, as they have options, which let them feel more secure (when they do not know the answer, they still can be the so-called "correct by default"). To check this, it is good practice to ask students to propose a supporting line.

By applying multiple-choice items to reading comprehension class the teacher can help the students arrive at meaning through inquiry (nature) instead of direct provision (nurture). Right, generally, context contributes greatly to the making of deductions about the meaning; nevertheless, control over the options put forward can activate the students' learned language. Imagine how three out of four options are assumed to have been familiar to students. Instead of saying "Well, the meaning of "X" is ..." (in the first language), it is better for a teacher to get the students to choose options which require reduced analysis. It would be possible that a well-

prepared exercise of this kind is viewed as a game or fun. The following are fairly good multiple-choice items for teaching reading:

I went to China for 3 weeks and it was an eye-opening experience. First of all the City of Beijing is simply huge with hundreds of tall apartment buildings in every direction. The roads and highways are wide and very clean with gardens and shrubs each side of the road. Traffic was not as chaotic as in many other large cities. The majority of cars and buses that I saw were new models, with very few cars older than a few years.

4. The word 'chaotic' in line 4 is closest in meaning to...

- | | |
|------------|-----------------|
| A. mistake | C. inexpensive |
| B. unknown | D. disorganized |

While in a good reading comprehension test a candidate is expected to make use of context (as clues) to lead him to an answer, in a reading comprehension class with quite rare clues for a difficult word a teacher can help the students by using a multiple-choice item with more 'identifiable' non-answer options. In relation with the example above, when the teacher assumes that clues are not enough and that the words in A, B, and C are most possibly detected by the students; he can expect much that the students will choose D as the answer. Though, it is not certain whether the students get right along with the understanding of meaning or not, hurt for meaning can go on afterwards.

Let's suppose the students have got only a very few clues, e.g. 'year' and 'to the old' from the text below, because the familiar words such as 'happy', 'wonderful', and 'warm' can be misleading (the words might give positive sense; and therefore reject option A), this multiple-choice item may be helpful in that the options proposed are all presumably recognized by the students (with reference to curriculum).



Happy New Year to all my friends

18. The word "adieu" in the text is most probably the same as...

- A. good bye
- B. greeting

- C. hope
- D. welcome

As is the case of matching exercises under condition no additional alternatives are given in the right hand column (considered bad for a test use), it is obviously good for introducing and developing vocabulary and a promising alternative to avoid translation technique in reading comprehension teaching (as many teachers do) in specific, and recourse to mother tongue in EFL teaching in general. Let's consider this illustration.

A

- | |
|--|
| <ol style="list-style-type: none">1. nationality2. diversity3. multiculturalism4. religion5. community6. approximately7. bilingual8. district |
|--|

B

- | |
|--|
| <ol style="list-style-type: none">a. roughlyb. a smaller part of townc. group of people (who share interest)d. citizenshipe. varietyf. many different culturesg. system of beliefh. being able to speak or write in two languages |
|--|

Suppose, a teacher of English has been thinking about the word "religion" as a potential problem to the students in a reading class, he can prepare and use the exercise above to introduce the word successfully under condition that the other words listed are assumed to be partly or fully recognized. This principle of "correct by default" – which is a weakness in testing – can be good in the use of multiple-choice for teaching. The way it works depends a lot on how the test writer manipulates the options.

Last but not least, because students are familiar with such exercise, anxiety when they take a test can be reduced. So, it is hard to deny that a teacher's good insight into and competency in language testing will be beneficial to the students. The knowledge about testing cannot only be good for its own sake, but also for learning.

Characteristics of Multiple-choice Items and the Consequence

As discussed above, in the context of language education, multiple-choice items type is really suitable for receptive skills training and testing because in receptive skills we deal with comprehension. It is,

A multiple-choice item usually consists of a stem (the part that is the question) and a number of options (ranging from 3 to 5). One of the options is the correct answer to the question and the others are called *distractors*. Here is an example:

(After a text of personal letter)

29. This letter was sent ...

- A. to Lizzy Ford
- B. from Canada

- C. by Stacy Ford
- D. on August 15, 2008

As we all might know already, to the teacher, one of the advantages of using multiple-choice items is that the response is simple to score. Similarly, to the students or test takers the advantage is that the response is short, just by selecting a correct answer and marking the corresponding letter A, B, C, D or E. But, it does not necessarily mean that deciding the answer is as easy as marking. With 4 options the chance of "true by default" is 25 per cent. Mathematically speaking, adding option will reduce the probability. But, it is just a little helpful, because with a poor multiple-choice item, the probability will rise accordingly. A wiser solution is to minimize the likelihood of guessing correctly to at most 20 per cent (with 5 options) by well preparing qualified items.

With those all characteristics, multiple-choice item is a seemingly too easy matter either in the eyes of students/test takers or teachers. To teachers, such perception may lead to serious threats. As a matter of fact, preparing this type of test item requires more considerations in addition to validity.

Whether a teacher develops, adapts or adopts instruments for testing, he frequently needs to take things into account. They include varying levels of the school or test population, content relevance, language use, and other features of the testing context, including culture. In the case of multiple-choice items, quality and all its consequences are often rooted in the options provided. One proof is that many canceled items are multiple-choice ones. Some teachers or good language learners rejected few items because no correct answers were provided.

It is necessary to note that in multiple-choice items it is not the answer that is multiple; but it is the option which is multiple. Therefore, in multiple-choice items the 'correct' answer must be genuinely correct. Sufficient literature, subject matter mastery, and item writing moderation

* The writer would rather not spend time discussing validity, because it is not confined to multiple-choice items type. It also applies to other test types.

(will be dealt with more fully later) is valuable to anticipate this problem. The following items have more than one correct or acceptable answer.

In cities __ (A1) __ poverty sometimes occurs in __ (A2) __ slums, ... People trapped in this kind of situation often have __ (A4) __ income.

44. A. much B. enough C. sufficient D. insufficient
E. inadequate

4. How do you blow the air into the balloon?
A. gently B. tightly C. quickly D. a little

Option which looks extreme might affect decision; the odd one out is likely to be a favorable choice, especially to any candidate in doubt about the answer, regardless of the correct answer. This is because the apparent appearance can attract attention. This condition, if true, is also illogical in that the odd one out can ignore the existence of stem and the other options within the same item. This obvious difference from the rest can be in size (usually the length) or content. According to the writer's experience, this is a very difficult step. Below is a good example:

(After a text on whales)

The following are the proofs that whales are not fish, except ...

- A. It has a single nostril C. The skin is smooth and shiny
B. It cannot survive on land D. Its tail consists of a pair of broad,
flat, horizontal

paddles

To : Hendry We invite you to attend our meeting. It will be held: - On Saturday, April 12, 2008 - At 8 p.m. - In the Village Hall Agenda: ...

When will the meeting be held?

- A. In the morning B. In the afternoon C. In the evening D. In the village hall

Another equally important problem deals with regularity of the keys that may result. By regularity here the writer means any pattern that provokes test audience to answer by guessing, unnecessarily specific

patterns. It can also be the constant repetition of the same key, e.g. a too long series of Cs. Once such regularity is recognized, decision can be interfered. It is easy for us to avoid this if we care enough. We can randomize or take another action in such a way that the keys are not predictably patterned. It is good practice, therefore, to pre-determine the list of keys (the letters) on a paper.

Another requirement with multiple-choice items is that each option should fit equally well into the stem. The test writer who wrote the following item did not seem aware that the indefinite article *a* had made option E fail to serve a full function, as a distracter. To at least some students the option becomes unattractive. They would not be likely to choose it because it does not fit into the stem: the indefinite article *a* cannot be used before a word beginning with a vowel.

People know that Ateng was a ... man. He was very good in playing a joke
A. killer B. usual C. contrast D. humorous E. indolent

Moderation

When we work alone, at times we do not see any limitations in our works. Everything seems right with them. Unfortunately, when dealing with multiple-choice items, the writer is familiar with items where more than one of the options is correct or, at least, acceptable. In this section we shall be discussing the value of showing our test item writing product to other people.

In the writer's experience, teachers are very reluctant to show outsiders the tests they have written. Wish they knew the significance of test item peer-reviewing. The rationales behind this are: (1) all kinds of tests should be valid; (2) the audience being tested is expected to have no assistance from others.

A colleague's knowledge may help. An individual teacher may have different levels of competence in the subject matter and/or testing. He might compensate for his lack of these by, for example, having his colleagues check the test items he has planned. And, you are lucky if a better experienced teacher is available at your workplace. It is highly advisable. Though when working alone you think an answer seems obvious, it is quite possible to write an answer that many colleagues would disagree with.

Test items have to stand on its own in that the students taking a good test have no possibility to get examples, demonstrations, or paraphrases for unclear parts. The test writer, therefore, has an obligation to ensure that all items are unambiguous prior to administration. Again, test item moderation can lend a hand.

While test preparation is underway, all those considerations elaborated above can free a multiple-choice test writer from taking a more time and energy-consuming procedures such as reliability and discrimination index analyses, which are strongly recommended for this

item type. Moreover, it is too far for a teacher to carry out such analyses, which are identical to calculations and mathematical analysis. Rather than try-out/pre-test the prepared test items to a group of students to get the data for the analyses why not, for example, exchange ideas with some better colleagues and/or friends in the local association of English teachers? Quite often, for example, a reading test is poor just because the candidates can answer the questions with no reference to the reading passage/s. In the case of multiple-choice items, the test writer may copy down the items without the reading passage and have friends volunteer to "take" the test. Those items that nearly 50% or more get right are probably poorly written. It's simple, and above all, likely to help promote academic atmosphere within your working environment.

The rest of this paper will be devoted to examples of good and bad multiple-choice items.

Poor

You will notice soon at least one thing wrong within each of the following multiple-choice items, partly taken from the national final examination.

(1)

Late in the afternoon, the boys put up their tent in the middle of the field. As soon as this was done, they cooked a meal over an open fire. ...

"As soon as this was done, ..."

What does the underlined word refer to?

- A. tent B. the afternoon C. the field D. the activity of putting up the tent

(2)

Vanessa Mae Vanakorn Nicholson is a musician. She is only nineteen years old, but she is a world-famous violinist. When she plays classical violin concertos on her electric violin, the world listens.

Vanessa Mae is Tai-Chinese. Her mother is Chinese and her father is Thai. She is from Singapore, but she now lives in London with her mother and her English stepfather. She's got cousins, aunts, uncles and grandparents in Thailand, Singapore, China and England.

What is the text about?

- A. A violinist B. A musician C. A pianist D. An artist

Dear Editor:

I am sick and tired of the traffic in this city! It is so bad that I can never get anywhere on time. There are many cars on the road, and most of them have only one person in them.

Another problem is the buses. They are so old and slow that nobody wants to take them. They are so noisy and very dirty. You can't even see out the windows!

Also the taxi drivers are rude. They never know where they were going, and they take a long time to get someplace. Taxis are expensive, too. And the subway is just too crowded and dangerous. What are we going to do?

George Grady is often late to somewhere because...

- A. the public transportation is rare
 B. the vehicles are very slow
 C. there are a lot of cars
 D. the traffic is bad

Good

Now, do you notice weaknesses such as those above in the following questions?

(1)

WOW...1000 Songs in Your Pocket!!!

If you love listening to music and if you like something small with an affordable price then having an MP3 player will be very handy. The digital audio player, often called the MP3 player, is very popular worldwide. This piece of technology is not something new. It's been around since 1998.

The good thing about having an MP3 player compared to a mobile phone or other audio player is that it comes in a range of colors and shapes. It's also very light and it fits perfectly in your pockets! They also have no moving parts so their batteries last a long time. You can jog or do other physical activities with them without causing skipping or damage.

Some MP3 players can store up to 40,000 songs. The more advanced designs of MP3s can store videos, data files, pictures and now there are even water proof MP3 players. This means that you can listen to music while swimming or having a bath!

34. Superiority of MP3 players over other audio players is discussed in paragraph/s...

- A. 1 and 2
 B. 2 and 3
 35. The word "handy" in the first paragraph is similar to...
 A. productive
 B. cheap
 C. 2
 D. 3
 C. useful
 D. safe

(2)

Kashmir is famous for its beautiful houseboats. Visitors to the Kashmir valley rarely leave without staying on one of these floating palaces. The houseboats float on the calm and jade-colored waters of Dal Lake, which has willow and chinar trees growing all around it. The houseboats are anchored among lotus fields and floating gardens.

Houseboats were first built by the British. When a Maharaja of Kashmir did not allow the British to buy and own land, they decided to build the houseboats and make them their homes.

Today the Kashmiris live in these houseboats. They have made nice hotels in the houseboats, too. Built out of cedar wood and decorated with a lot of care, the houseboats are showpieces. They normally range from 24 to 38 meters in length, and three to six meters in width. A houseboat has a living room a dining room, and two or three bedrooms with attached baths and hot and cold running water. The rooms are built and decorated imaginatively to make guests' visits as comfortable and pleasant as possible.

39. The main idea of paragraph 2 is...
 A. the history of houseboats of Kashmir
 B. the size of houseboats of Kashmir
 C. the beauty of houseboats of Kashmir
 D. the fight of Kashmiris against the British
40. The following are right according to the text except...
 A. To visitors, staying on the houseboats is a good experience
 B. The houseboats of Kashmir float in the open sea
 C. The British do not live in the houseboats anymore
 D. The Kashmiris have decorated the houseboats

(3)

Football is the most popular sport played and watched in Indonesia. But now FUTSAL, which has been played in Indonesia since 2001, is becoming more popular than football. FUTSAL is an indoor game originally from Uruguay. The term FUTSAL is the international term used for the game. It is taken from the Spanish or Portuguese word for soccer, FUTbol or FUTebol, and the French or Spanish word for indoor, SALon or SALa.

FUTSAL is so popular because it (1) is safer than football because there are less collisions with other players; (2) needs a fewer number of players (5 in a team) than football; (3) needs a small place to play in; (4) has much simpler rules; (5) teaches better skill development.

FUTSAL is played with a smaller ball with less bounce than a football. Even though the ball may go out of bounce, the ball must be put back in play within four seconds or the opposing team gets possession. This not only encourages better control but it also keeps players moving. The emphasis is on improvisation, creativity and technique. It demands quick reflexes.

12. These are the reasons why futsal is more popular than football in Indonesia except...
- | | |
|--|-----------------------------------|
| A. it needs small space famous players | C. it is played by famous players |
| B. it uses simpler rules fewer players | D. it is played by fewer players |
14. The word "This" in paragraph 3 refers to...
- | | |
|-----------|---------------|
| A. Futsal | C. The rule |
| B. Bounce | D. Possession |

(4)

Races in the Badwater Run in California ...⁽⁴⁰⁾... 139 miles (224 kilometers), climbing 8,653 feet (2,637 meters). The race begins in Death Valley, a desert. The Temperature is about 130⁰F (54⁰C), and ...⁽⁴¹⁾... sometimes run through sandstorms. The race ends near the top of Mount Whitney, ...⁽⁴²⁾... the temperature is only 30⁰F (-1⁰C), and there ...⁽⁴³⁾... sometimes ice storms. ...⁽⁴⁴⁾... winners can usually finish in about 28 hours.

- | | |
|----------------|--------------|
| 40. A. run | C. ran |
| B. runs | D. will run |
| 41. A. winners | C. teams |
| B. players | D. |
| contestants | |
| 42. A. whose | C. who |
| B. where | D. which |
| 43. A. was | C. are |
| B. will be | D. be |
| 44. A. Amazing | C. Amazingly |
| B. Amazed | D. Amazement |

(5)

ROYAL PLUS
Insulated Glass Vacuum Bottle
KEEPS IT HOT OR COLD

PRE-CONDITION THE FLASK
BEFORE FILLING
FOR OPTIMUM RESULTS, FILL TO
CAPACITY

EACH BOROSILICATE GLASS
FLASK IS TESTED BEYOND
INTERNATIONAL STANDARDS FOR
BETTER PERFORMANCE

6. The product discussed in the advertisement is a

A.



C.

B.



D.



(6)

Maura, who like to be thought of as the most beautiful and powerful queen of Arabia, had many suitors. One by one she discarded them, until her list was reduced to just three sheiks. The three sheiks were all equally young and handsome. They were also rich and strong. It was very hard to decide who would be the best of them.

One evening, Maura disguised herself and went to the camp of the three sheiks. As they were about to have dinner, Maura asked them for something to eat. The first gave her some left over food. The second Shek

gave her some unappetizing camel's tail. The third sheik, who was called Hakim, offered her some of the most tender and tasty meat. After dinner, the disguised queen left the sheik's camp. The following day, the queen invited the three sheiks to dinner at her palace. She ordered her servant to give each one exactly what they had given her the evening before. Hakim, who received a plate of delicious meat, refused to eat it if the other two sheiks could not share it with him.

8. How did Maura come to the camp of the three sheiks?

- A. By changing appearance with the sheiks
 B. By setting up a camp nearby her servant to spy them
 C. By having a date
 D. By asking

(7)

(heard from recording)

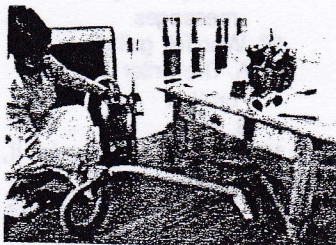
- Man : May I watch what you're doing?
 Woman : Sure. You dig a hole, put in the seed, cover it with dirt, and then water it.
 Third voice : Which one is the woman in the conversation?

(seen on the test sheet)

B



D



ask Toni



Dear Toni,

Someone told me that my brother's girlfriend was dating another guy. I felt I should let my brother know, and after I did, he confronted her with the story. Although she denied it, it caused a terrible argument and they almost broke up. Now it turns out that the rumor wasn't true, and my brother has stopped speaking to me.

Distraught Sister

Dear,...

Well, you learned a lesson. You shouldn't have listened to gossip. And you shouldn't have passed it on. Now you have to repair the damage. Apologize sincerely and hope that he will forgive and forget.

Toni

16. This kind of text can most probably appear ...
- A. on TV screens
magazines
- B. on greeting cards
messages
- C. in
- D. in short
17. Who felt guilty?
- A. The writer
brother
- B. Toni
- C. The writer's
- D. Another guy
18. "Although she denied it, it caused a terrible argument and they almost broke up."
The underlined phrase is synonymous with...
- A. got angry
- B. separated
- C. had relationship
- D. married

Conclusion

Instead of its limitations, multiple-choice items can be considered as the most advantageous item type for either testing or teaching receptive language skills, which have nowadays been emphasized in Indonesia. Therefore, to help their students achieve considerable success the teachers should improve and maintain their competency in evaluating, preparing, and using this type of test items both for classroom exercises and testing.

References

- Alderson, J.C. et. al., 1995, *Language Test Construction and Evaluation*, Cambridge: Cambridge University Press
- Brown, D.J., 1996, *Testing in Language Program*, New Jersey: Prentice-Hall Regents
- Hughes, Arthur, 1989, *Testing for Language Teachers*, Cambridge: Cambridge University Press
- Madsen, H.S., 1983, *Techniques in Testing*, Oxford: Oxford University Press
- Suwartono, 2007, "A Good Teacher of EFL Speaks the Language in the Class", *Human Resources Development in English Language Teaching* (The 55th TEFLIN International Conference Proceedings)
- Suwartono, 2008, "Dependent upon Teacher's Creativity", *Globalization and Localization in CALL*, (GloCALL 2008-Indonesia)
- Weir, C.J., 1990, *Communicative Language Testing*, New York: Prentice-Hall