Lessons Learnt from Vetting English Examination Papers

MAYA KHEMLANI-DAVID
University of Malaya

In many a teaching institution, teachers apart from teaching, are given the responsibility of setting examinations. Generally a teacher is made responsible for the preparation, editing and typing of her own paper.

We, in the University of Malaya, have "indulged" in vetting sessions. This means that after the paper is set by an individual or two, it is then reviewed by other members of the teaching team. Members will then make comments regarding the nature, type, wording, format etc. of the questions and these suggestions, if accepted, will have to be incorporated in a revamped paper. Provided the suggestions are not to belittle the writer or antagonize him, and if the suggestions are accepted in the same spirit, that is, that the ultimate goal is~ a good product, a valid and reliable test, the vetting session can result in a lot of learning of good habits and unlearning of poor test habits even for the most experienced of teachers.

It is the objective of this paper to record the kind of errors made, which were pointed out at the vetting sessions and consequently to set forth some general guidelines for the setting of ESP test papers. These guidelines resulted from the discussions during the vetting sessions.

Background to the Learners and Value of Tests

The objective of both the instructional and examination phases is the attainment of reading skills, both at linear and nonlinear levels. The test paper, thus tests reading skills with Sections A, B, C focusing on linear tests while Section D concentrates on nonlinear texts eg. graphs, charts etc.

Since writing skills are not taught it would not be fair or valid to test them. Consequently, the questions take a multiple-choice format and the student is merely required to circle the correct answer/key.

The texts themselves, both linear and nonlinear, have to be authentic and ESP based since the focus at the teaching phase was based on authentic and ESP based materials (in this case Economic texts). The texts for the examination are generally extracted from recent economic journals although the questions are devised by the test-writer. Test-writers are permitted to "doctor" the texts so that reading strategies like the use of reference markers and other discourse linkers stand out and can

become 'testable' items. Such doctoring of the text is in itself fraught with dangers because incorrect plural/temporal markers which do not flow from the earlier text may be inserted - a consequence of doctoring. It is important therefore that when doctoring parts of a text that the test-writer and vetters should read it from the beginning to ensure that there is a flow of thought and language which is in line with the new revised doctored text.

Choice of the Text

The texts selected, apart from being ESP based, must be of a level commensurate with the students' ability and moreover must lend themselves to the setting of a number of questions for example 10 from text A in Section A. The text should permit a range from questions - of inference, prediction, factual, implication etc. If the text is very factual and has no development of ideas then it will be rather difficult to set a variety of questions from the text. Consequently, in an initial vetting phase, even before the questions are written, the chosen texts are vetted by the team of teachers. The text finally selected must lend itself to asking of different types of questions. Additionally, the language content of the text must be such that even though it may be lexically difficult there should be enough clues or signals - language markers which can be used to decode the text.

The person who has selected the texts should indicate roughly at which point a question could be asked and the nature of the question eg. factual, inference, summary etc. Vetters then meet to determine if the text is suitable ie. it is ESP based, is in line with cultural and value norms in an Islamic country, does not contravene any regulations or laws, does not upset any racial group and is sensitive to racial innuendos, does not paint a poor picture of the leading political figures etc. If content-wise it is acceptable then the number and type of questions it is possible to extract from the text are then scrutinized. There should not be too many factual questions or summary questions for instance. A proper mix of the various skills earlier taught should be tested.

The texts selected for year I and year II students should be compared so that they reflect a range of difficulty. However, at times even though a text may appear primae facie, more difficult/implicit rather than explicit this can be overcome by simpler and more straight-forward questions and on the other hand a simple text in terms of vocabulary level, may lend itself to a more complex level of questions eg. inference and predictive questions.

The lexical context, that is vocabulary level of the text, must also be reviewed. Are there too many difficult words? Can complex and difficult words be decoded? Are there enough cues in the text to help the test-taker decode the mystery of' the problematic lexical item? Are there sufficient discourse markers and reference markers and the like which could help the test-taker decode the text?

The length of the text should also be considered. It should not be too long. The general consensus was that a suitable text of about 600 words could lend itself to a range of about 10 questions.

Besides Reading Comprehension - RC (Section A), linkers are tested in Section B. The broken text selected for this must lend itself to testing such connecters. There should be enough cues to help the test-taker key into the correct answer. Sometimes the vetting team felt it was easier to change a text rather than to attempt to salvage a text by doctoring it.

When transcribing from the original text it is best simply to make a copy of the original. Rewriting text can in themselves cause problems as is evident from the following example. We spent numerous hours of argument over the following phrase:

"numbers of the poor"

The use of the article "the" was questioned as incorrect and inappropriate. However, the test-writer was not happy about changing the item as it "was authentic". Text-writers must however not be afraid to correct the original authentic text if there are grammatical errors since text writers/article writers are not infallible. In this example, however, the original writer was not at fault. The article "the" had appeared only in the rewriting of the text and on checking with the original text it was not found. Consequently, the test-writer was willing to make the change to align it with the original text - as it was.

Types of Questions

Of the 10 - 15 questions in Section A there should be a couple of summary type questions for example:

"The most suitable title for the text/passage is:-

One or two questions should test result: a couple should test main points and a couple should test vocabulary.

However, in the Reading Comprehension Section, such vocabulary items should be indirectly tested eg. Why are changes "traumatic" (line 15) for European farmers? Indirectly, the question tests an understanding of the lexical item 'traumatic'.

There could also be a couple of true/not true questions, for example:-

Which of the following is NOT True of East European farmers?

Level of Difficulty and Type of Questions

The level of difficulty depends on the level of proficiency of the students. If they are of the lower level than direct and straight forward questions should be asked rather than questions on inferences/implications. There should be less inference questions and more literal questions for students with lower proficiency.

Additionally, the loading of distractors in the options given should be reduced for such students and shorter distractors should be provided.

Questions must be so worded that they are crystal clear to the test-taker. He must not be unable to answer the question simply because he does not understand the question. The student must therefore not be confused by the question itself. A question like

"The results of the recent election showed....."

confuses the student as he may not know exactly what is required of him since the question is so open-ended. If the structure and phraseology of the question is changed to that of a direct question the testing point would be clear to the test-taker. An example of a direct question would be

"What is the result of the recent election?"

Testing Vocabulary Items

When selecting vocabulary items to be tested it is imperative that the items selected for testing should have, in the text, sufficient contextual cues for the test-taker to make an educated guess.

Incidentally, when referring to items lifted from the text it is vital to give the line reference and also underline the word in the text so as to make it clear for the student since it is possible that the same vocabulary item may appear more than once in the text and have a different meaning in both instances.

When selecting vocabulary items ensure that if the item tested is a phrase it should not be cut/chopped up. "Pour out" was tested yet in the text it was "pour out of'. The vetting team was of the view that the whole phrase should be tested.

Testing Linear and Non-Linear Knowledge

Section D of the test paper evaluates the ability to match linear with non-linear text. At times a number of texts and diagrams are given and test-takers are to make the correct match. Initially, test writers gave four texts and four diagrams but the chance score with 1:1 is also 1:1. Consequently, the number of texts or diagrams had to be increased so that the chance score was reduced and the test then became not merely one of chance/luck/guess work. At the same time, the texts which have been extracted should include notions and semantic themes which help the learner to cue in to the relevant diagrams. It would not be fair to the test-taker if' there were no cues between texts and diagrams. On the other hand, too many cues would make the test item too simple and it would not be testing anything.

Testing of Linkers

If linkers are to be tested it is important that they should not be confused with relative pronouns (like when) or dangling modifiers like "with orders declining". Although "with" can be a linker it is imperative that the test writer - checks the context, for in the text "with" may be acting not as a linker but as a modifier of sorts as in the above example ("with orders declining").

Language Used in Options

In the writing of MCQ it is important that the options themselves are understood. The vocabulary used in the options must be simple and understood by the test-taker. The vetting team simplified the following lexical items in some of the options

```
enmassed     ---------->collected

optimistic   ---------->hopeful
```

It must be remembered that the test is aimed at measuring the understanding of the prescribed text. The test should not therefore, test the understanding of the question itself. If the question or lexical items in the question are not understood then it is not a fair test of the text.

The language used in the options should not be colloquial. Test writers should be wary of injecting oral and or colloquial language in options. The following example is self-explanatory.

1. First draft by test-writer:-
   ' They will be getting back their ancestral land'.
2. Second draft after vetting session:-
   'They will be returned their ancestral land.'
3. Third and final draft after vetting:-
   ' Their ancestral land will be returned.'

Even the tense used in the option should be consistent. Not to confuse the test-takers and also to have an internal structural consistency, if one option has a past tense marker, all options should also be in the same tense.

The Writing of Options in a MCQ Test

In Section B where knowledge of linkers is tested it is imperative that distracters must be of the same class as the key i.e. that all options must also belong to the linker class. In one question the four options initially presented were: even more, on the other hand, moreover and too. The distracter "too" was glaringly, out of place.

Even when the options are linkers, it is important that they indicate the same degree of semantic meaning. One question held the following options, "thereby", above all, better still and without". The last linker did not show a movement of thought as did the others and was disqualified.

The test writer must be vigilant and not provide two options which are possible keys. For example, when a key is an additive linker like "as well as" it is important the other options should not include other additive linkers, which could also be possible keys, for example "in addition".

It is also not advisable to have three negatives and one positive option as in the following:- "rise, decline, came to a halt, fluctuate", as the negative "decline" stands out, It could either be the key or a distractor. In options it is important that all options are approximates rather than opposites. The chance score is reduced it there are three approximates and one completely divergent option. This kind of distracter is against the grain of testing.

Options which are similar should also be avoided as in, "consecutive, continuous" since they could both be possible keys or if they are distracters if one is eliminated the learner will also automatically eliminate the other.

All options written must involve some processing of text on the part of the test-taker. Sometimes an obviously wrong answer does not require any rechecking with the text for example "the poor employed in the rural non-farm sectors are lowly paid". Alternatively, the key must not be so obvious that it is a matter of logic rather than processing of the text as in the following example:

> Which of the following is NOT a feature of poverty?

> (A) The poor are mainly to be found in regions where resources are plentiful.

It is imperative that one option should not stand out. During one vetting session, for example, we discovered a question where three of the four options had a percentage whilst the key did not have a percentage. All students chose the latter option and the item did not discriminate between the poorer/better students as it was what students call a 'gift question' but setters call 'a goof.

The same word should not appear in two distractors as in the following examples:

a. take long leave
b. leave in big numbers

as it reduces the chance ratio.

Keys

Keys or the correct option have to be clearly the correct answer. In one question there were two close keys. It was therefore necessary to doctor the text so that it was obviously one rather than the other. The question read "The main difference is.... yet two options were possibly correct since the text did not indicate the difference as the main one. Consequently, the text was doctored to include the word above all so that there could only be one key as a result of this inclusion.

The key must not be so obvious since if the first option is very obviously the key, the student does not need to read the text further not does he even need to read the other options.

Standardization of Language in the Questions

The rubrics must be standardized and consistent. Test-writers when setting questions tend to concentrate on the nature of the question and the options rather than the consistency and standardization of the rubrics.

Section B of this particular test paper tested decoding of reference markers. One test writer started the question by

> "Their" in line 20 refers to:-

while another test-writer phrased it thus

      In line 20 "their" refers to:-

After some discussion by the vetting team the latter format was selected and test-writers were informed that this format would be consistently used in the future.

Spelling

Certain policy rules have to predetermined before a paper is to be written. Should the spelling follow American or British orthographic traditions? We were of the view that it was important for our learners to be aware that divergent spelling systems existed, consequently, a decision was made that when a source was American then the spellings would be American and similarly with the British texts.

Punctuation

Even punctuation has to be checked. As language teachers it is imperative that we set ourselves as good models. Even the most experienced of teachers can make errors as witnessed in the following example:-

```
*cooperative --------------> co-operative

*four year period -------------->four-year period
```

Conclusion

The importance of vetting and revetting examination papers cannot be overstated. Many variables are involved in writing an exam paper starting with

a. the choice of an appropriate text;
b. the nature and range of questions tested;
c. the nature of options and language used in options;
d. rubrics/standardization and
e. punctuation and spelling norms.

No man is an island and so is no English teacher especially where the writing and preparation of an English test paper is at stake.