

A Rough Guide to Item Writing: Multiple Choice Test Design for Beginners

By Graham Barrow of GR Business Process Solutions

Over the past months I have written about the role of analysis in multiple choice tests, the effects of guessing in test scores and why test scores are not always reflected in sales quality. In this article I would like to turn my attention to a much more fundamental skill – producing good quality test items from scratch.

With the imminent advent of regulation within the General Insurance and Mortgage markets many more firms are going to have to demonstrate the competence of their sales forces and many of these firms will not previously have tested their sales force. Larger firms will have access to the expertise and technical know-how to embark on this project, but for a lot of the smaller firms (who make up the majority of the soon-to-be regulated high street firms) this is not quite such a straightforward undertaking.

Nevertheless, the FSA will expect them to approach testing in a robust and effective manner so I thought it might be helpful for those who will soon have to focus on this task if I were to outline some basic ground rules and techniques for the production of a cost-effective and reliable testing environment.

The starting point must be the syllabus. You simply cannot produce a test that has any meaning unless you can point to a corpus of knowledge that the candidates sitting the test must know, in order to demonstrate competence in their chosen field. Ideally the syllabus should be hierarchical, starting with high level subject areas and becoming more detailed at subsequent levels. The relevant appendix to CP157 (available from the FSA website) provides an excellent example of syllabus construction.

A good syllabus will also define the level of knowledge required for each syllabus area. For example, the candidate might need to *know* minimum premium levels, *understand* how a policy works, and be able to *apply* that knowledge to a given situation. If the syllabus clearly states these requirements, writing suitable items to test them becomes much easier.

So, once a comprehensive and clearly written syllabus is in place, what then?

Well, it is unlikely that every area of the syllabus will have equal weighting, so it is now important to ascertain what proportion of each syllabus area should be represented in any test. You may decide, for example, that knowledge of policy conditions (such as premium levels, minimum and maximum ages at entry etc.) might constitute 20% of the overall test, demonstrating a clear understanding of how the various policies work should constitute 40% of the test and the remaining 40% should demonstrate the candidates' ability to apply that knowledge to given situations. It's also worth point out that many firms do not bother with the knowledge elements at all, because to demonstrate understanding and the ability to apply that understanding, you must have the underpinning knowledge in the first place and, if you don't, you can always look it up.

At first glance that may appear to fly in the face of the point of testing in the first place. Not necessarily. I am in the middle of having some fairly major building work done to my house. I don't expect the builders to know every type of brick, insulation material or bath fitting currently available. On the other hand, I do expect them to understand how they all fit together and be able to do so competently. I also expect my doctor to be able to understand any symptoms I may have and work out what is wrong with me. I don't expect him to know every drug available and all the potential side effects related to each drug.

So you see knowledge is something you can always look up, but you cannot do this with comprehension and application skills; these **have** to be learned.

Now we know what the subject matter is and what type of questions we want to ask. What other rules should we have in place in order to ensure the effectiveness of our items?

Right at the top of the list I would place the need for clear and simple English. Nothing frustrates candidates in an exam more than questions that contain convoluted clauses and exotic vocabulary. I love words and enjoy discovering new ones, particularly when they describe something or evoke an emotion that, hitherto, had required many. Just this week I have had occasion to use '*rapprochement*' and '*palimpsest*'. These are lovely words that, in the correct situation, have no replacement. However, it is highly unlikely that they would (or should) appear in the average multiple choice test as their usage is more likely to confuse than enlighten the candidate.

Next is the need to ensure that the question is framed using clear and straightforward grammar. How many people would confidently answer a question constructed as follows: -

Q. When would it not be true that, if you were not accepted for underwriting, a policy could not be put on risk?

Inclusion of successive negatives makes understanding more and more difficult until the point is reached when the candidate has no idea what is being asked of him.

It's also very important to leave out any unnecessary words. Consider the following question: -

Q. Jose Raul Capablanca was born in Havana, Cuba in 1888. He learned to play chess at the age of 4 and won the Cuban Chess Championships at the age of 12. At what age did he win the Chess World Championships?

If the point of the exercise is to test the candidate's knowledge rather than demonstrate the item writer's, this question would be much better written as follows: -

Q. At what age did Jose Capablanca win the Chess World Championships?

Next on the list is to ensure that your question is only testing one learning point. Why? Because if the candidate gets the question wrong you cannot be sure which element (or elements) they did not know. Sometimes, if they get it right, it will not be possible to tell whether they knew both learning points, or only one. Consider this question:

Q. Who scored West Ham's last goal in their most recent FA Cup final, and in which year?

- a. Trevor Brooking in 1980*
- b. Gerry Francis in 1982*
- c. Alan Sunderland in 1983*
- d. Ian Rush in 1985*

Clearly, because of the options given, it is necessary to know **ONLY** the year **OR** the scorer. As the setter of the test, though, you won't know which one the candidate knows, or whether in fact he does actually know both.

Finally a stem should either be a direct question (as above) or should form the first part of a statement, for example: -

Q. The current basic rate of Income tax is...

OK. That's covered some of the basic rules of writing stems; now let's look at the options (both the correct answer and the distractors). Here again there are some key rules to learn.

First, and most importantly, the distractors should all be both plausible and attractive. There is no point having a four option test item that contains one correct answer, two potentially correct answers and one non-hoper. You might as well just have three options in the first place. Let me bring this into focus by asking the same question with two different sets of answers and you will see what I mean.

Q1. Vera Baird is the sitting Labour MP for...

- a) Greencar*
- b) Bluecar*
- c) Redcar*
- d) Yellowcar*

Q2. Jacqueline Smith is the sitting Labour MP for...

- a) Redditch*
- b) Reigate*
- c) Rochdale*
- d) Rotherham*

I would be prepared to bet a lot of money that the vast majority would get the first question right but struggle over the second. I realise this example is somewhat extreme (although, sad to say, I have seen similar examples in real tests) but the point will have been made. If you think about it, the first example would be suitable for the £100 question in "Who Wants to be a Millionaire" but the latter would be tough, even for the £1,000,000 question, and yet the only real difference between the two is the distractors.

I can't emphasise enough that the quality of any question is governed by the quality of the distractors. There is the world of difference in candidates being able to answer a question by eliminating three wrong answers than by positively selecting the correct one. Knowing what is wrong is not the same as knowing what is right.

Once you have selected three plausible distractors to go with your correct answer you now need to ensure that they are roughly the same length. You may wonder why, but it is one of the unwritten rules in answering multiple choice questions that, in the absence of knowing the correct answer, candidates tend always go for the longest, and there is a very good reason why this might work.

Anyone who has written items in the past will testify to the fact that producing good quality items is a fairly long and difficult process. It is clearly very important that the correct answer should not be open to doubt so great care tends to go into constructing this option so that all possible bases have been covered. However, the same care does not tend to go into the distractors and this gives rise to rather lopsided answers. Consider the following example: -

Q. What restrictions exist to (i) investments and (ii) benefits, for the Protected Rights element of a transfer value into a Personal Pension Transfer option?

- a) None for either*
- b) None for benefits, with-profits insisted upon for PRP*
- c) With-profits for PRP, TFC restriction for benefits*
- d) Benefits cannot be taken before age 60. Benefits must be taken as pension rather than TFC. Pension must be purchased on unisex annuity rates. Indexation must apply. No investment restrictions.*

This question manages to commit three sins in one go! Firstly, it is testing more than one learning point so, as explained above, if the candidate gets the question wrong it is not possible to ascertain whether he knew part of the answer or none at all. Secondly, the question is littered with jargon which could disadvantage a candidate who knew the underpinning knowledge but not the three letter abbreviations which form part of the options offered. Thirdly, option “d” is both significantly longer and much more detailed than any of the other three options and is strongly clued as being the correct answer.

There is, of course, a great deal more to writing items than this, but adherence to these simple rules will without doubt give rise to a robust and reliable question bank.