or by, for example, the 'Chair, EFL officer, independent vetter'.

QUESTION 20: *What typically happens to the draft examination after the above committee has deliberated?*

The most informative way to report the answers to this question is to list examples of the different procedures:

1. Manuscript to printer, artwork commissioned, tapes recorded; proofs circulated to Chief Examiners and Moderator, worked through by scrutineer; final proof passed for press with print order.

2. Items approved subject to agreed amendments. Preparation and proof-reading of 'vetted' copy. Return to Chief/Senior Examiner for a final check and approval to ensure that examination has been prepared in accordance with vetted and agreed copy.

3. Once an agreed version of a paper has been completed it is (computer) typed and submitted for printing. All members of the revision committee see the first proof and are given the chance to make alterations (major and minor).

4. [Exam Officer] revises draft in light of reports from five senior examiners and arranges for production. During recording, actors comment on clarity and naturalness of language. Final text and copy of tape sent to Director who arranges printing of exam texts and copying of master tape.

5. Chair of paper and subject officer put together two parallel versions.

3.8 Survey of EFL Examination Boards: Documentation

The only documents which expanded on the answers given in the questionnaire related to the construction of test items. City and Guilds sent us two of their booklets, *Setting Multiple-Choice Tests* (1984), and *Setting and Moderating Written Question Papers (Other than Multiple-Choice)* (undated). The former gives helpful ideas on how to construct multiple-choice questions, cites a wide range of examples and advises writers about some of the potential pitfalls. The latter gives advice about rubrics and the layout of non-multiple-choice questions, and accompanies recommendations on how to design good items with examples of poor and improved questions.

Pitman sent us copies of their 'Blueprints' for each level of their English for Speakers of Other Languages (ESOL) exams. These are guidelines for item writers, and not only describe the kind and level of language to be tested but also give instructions about text types and advice about how to write good items.

3.9 Discussion

As can be seen from the answers to the above questions, most of the examination boards treat the item writing-process seriously. They give item writers ample time to produce future papers, and carry out thorough checks on the draft papers.

The one area which does not always get sufficient attention relates to coverage of the syllabus. Although almost all the boards tell their item writers that their test papers must cover the syllabus, only half the boards check that the papers actually do. Since some areas of a syllabus are always easier to test than others, item writers sometimes find they are unable or unwilling to test the more difficult aspects, and, because of this, the content of some test papers may be unbalanced. We feel, therefore, that it is essential to check draft papers to see that the syllabus has been covered adequately.

3.10 Checklist

1. In order to understand what a test item is doing, it is essential that you respond to the item as a test taker. Simply 'eyeballing' the item is quite insufficient.

2. Taking your own items is important but inadequate – you 'know' what you think the item requires. Therefore, get others who have not written the item to take it, as a test taker would.

3. Nobody writes good test items alone. Even professional test writers need the insight of other people into what they have produced. Therefore, get other people (plural!) to take (i.e. respond to, not just look at) your items.

4. Don't be too defensive about your items: be prepared to change or drop them if others see them as too problematic. We all – really, all – write bad items.

5. Get respondents to say why they gave the answer they did, and if possible to say how they went about answering the item.

6. Again, if at all possible, get your respondents to say/write what

they think the item is testing, independently of anything you may think it is testing. In other words, do not tell them what you think it is testing and then ask them to agree! Also, ask respondents to say what they think the main purpose of the item is, and what level of student it is suitable for.

7. All tests should be edited or moderated by people who have not written them. This editing committee should have available the responses of your respondents at some point in their deliberations. Items that have provoked unexpected responses from respondents must be revised.

8. If there is a defined test population, get respondents or editors to estimate roughly what proportion of the test population will get the item correct.

9. Match what most or all respondents agree the item is testing against what the test writer says it is testing. Resolve disagreements.

10. Match the agreement in 9 above against the test specifications or the syllabus.

11. Look at the syllabus or specifications and ask yourself if anything significant is missing from your test. If so, is this justifiable?

12. Ask yourself if the test method will be familiar to students. If not, change the method, or ensure that the instructions are clear. Ask yourself if another test method might be more suitable to your purpose, or clearer/easier for candidates.

13. Ask yourself what the item/collection of items will tell you about the learners' abilities. If the test/item results disagree with your opinion of the students, which will you believe – the test or yourself?

14. What are the chances of the students getting the same result if they took the test again the next day?

15. Pretest the test on students who are as similar as possible to the target students. Examine their responses and ask yourself:

   i) Are there any unexpected responses? If so, are any of them unexpectedly correct? If yes, add them to the mark scheme or change the item.

   ii) How many students found an item easy? Is it too difficult or too easy?

   iii) Which students got the item right – the stronger students or the weaker ones? In theory, the stronger ones ought to do better on each item, but in practice the item may contain a trick, an obscurity, two correct answers, or some such problem.

16. Get respondents or students to take reading and listening comprehension items without the associated (written or spoken) text. Can they still get the item right? If yes, it's probably not testing comprehension of the text.

17. With listening comprehension items, ensure that respondents actually listen to the text (rather than read it) when they respond to the item. Reading the text is easier than listening – they can do it in their own time, pause and re-read, and so on.

18. Is the language of the item easier than the language of the text? If not, you are also testing understanding of the items.

19. In multiple-choice questions, are some distractors possible – in some standard variety of the language; with a different interpretation of context; with different stress and/or intonation? Is the correct answer obvious because of length or degree of detail?

20. Have all possible/plausible answers been foreseen in the answer key?

21. Is the item contextualised? Is the context sufficient to rule out other alternative interpretations or possible ambiguities?

22. Is the item likely to be biased against or in favour of students of a particular gender/culture/background knowledge/interest?

23. How life-like is the item? Does the item look like something that students might have to do with language in the real world? For example, with writing tasks, do students have a purpose for writing and someone to write it to?

24. Would it be desirable to present the instructions, or indeed the test items, in the mother tongue?

25. How will the performance of the candidate be judged? Are the marking criteria or the correct or expected responses specified? Are they specifiable, or must you wait until you have a few sample performances before you can finalise your mark scheme?

Bibliography

Alderson, J.C. 1978. *A Study of the Cloze Procedure with Native and Non-Native Speakers of English*. Unpublished Ph.D. thesis, Edinburgh University.

Alderson, J.C. 1979. The Cloze Procedure and Proficiency in English as a Foreign Language. *TESOL Quarterly* 13 (2): 219–227. (reprinted in J.W. Oller, 1983. (ed.), *Issues in Language Testing Research*. Rowley, Mass.: Newbury House.

Alderson, J.C. and D. Wall. 1993. Does Washback Exist? *Applied Linguistics* 14: 115–129.

Allan, A. 1992. Development and Validation of a Scale to Measure Test-Wiseness in EFL/ESL Reading Test Takers. *Language Testing* 9: 101–123.

Buck, G. 1989. Written Tests of Pronunciation: Do They Work? *English Language Teaching Journal* 41: 50–56.