NEW MULTIPLE CHOICE FORMATS: THE OMISSION ITEM

HANSPETER BAUER AND HELMUT BONHEIM

UNIVERSITY OF COLOGNE

As the general level of competence in English rises all over Europe, the simpler forms of measuring that competence are increasingly inadequate. The old staples of multiple choice tests were vocabulary, grammar and reading comprehension, but increasingly we need to measure higher-order competencies. Most of our graduates who train to be teachers in fact take up other careers: a recent count at an English and a German university showed that only 9% (the figure happened to be the same in each estimate) of the graduates go into teaching. The others go into commerce, industry, publishing, administration, advertising, politics, where they need to write, edit, correct and revise texts, make them more effective, not simply more correct. How do we test such forms of competence?

The omission item poses the sort of problem with which the graduate is often faced: what word fits where, which ones are essential, which are superfluous? This is how it looks:

1. Which word or set of words can be omitted from the sentence?

Remember that $\underbrace{\text{there is}}_{A}$ nothing in $\underbrace{\text{the}}_{B}$ nature of our laws which B forbids a citizen from building houses as high $\underbrace{\text{as the Eiffel Tower}}_{C}$

Miscelánea: A Journal of English and American Studies 17 (1996): 63-70.

and circular in shape.

The stylist will seize on "there is"—surely it is superfluous? "Remember that nothing in..." sounds more decisive and efficient as the opening of a sentence. But then comes the "which"—not at the candidate's disposal here, and thus not to be thrown out. The "which" construction forbids the omission of the word group marked "A" after all. As to "the nature"—this is a trap for the French or German student, for in these languages *nature* often requires a definite article. But here "the nature" is *noli me tangere*. Option C, *as the Eiffel Tower*, could be omitted. But then we would have to carry out further changes in the sentence, like excising the "as" which precedes "high." So the key must be D, "in shape." This is an addition which, although it does not violate laws of grammar or of idiom, is simply "deadwood," a tautology which needs to be cut if the sentence is to meet high standards of good English.

It will be immediately apparent that this is potentially a "discrete point" item, using the principle of the "sore finger" format (Bonheim and Kreifelts 1979, Bauer 1991). One might think it assignable to the general area of "idiom." But further considerations of various kinds, not only of vocabulary or idiom, will play a role if the candidate is to come to the one and only correct answer. The sentence as it stands is grammatical, though not elegant. But that is not the point here. The item constitutes a test of editing skills, and needs to be solved at the level of what is nowadays called "text grammar."

For over twenty years we have been using the omission format (we call them "omis," a pun that plays on the diminutive which young German children use for their grandmothers, the "Omas"). They have been included in university entrance examinations and in national scholarship tests. The point-biserial discrimination indices are on average almost double those achieved in standard items involving English idiom. Apparently the right solution to an omi is often based on an application of sentence logic rather than on the application of discrete-point language skills. Thus the candidate who can do one omi correctly is likely to be good at doing the others as well.

The reason for this becomes evident if we look at the conclusions which the candidate must draw so as to find the right answer to the item cited above. The interrelations of a number of words and phrases have to be taken into consideration, partly on a grammatical, partly on an idiomatic or stylistic basis. An overriding sentence logic plays a role as well. A student who has all the skills needed to construct correct and acceptable sentences and texts will

have fewer problems in finding the right answer than will the candidate who focuses on only one point at a time.

A little more difficult is the kind of omi in which the one *essential* element in the sentence has to be located. Our first and second examples are simple grammar items:

2. Find the word which is essential to the sentence and cannot be omitted:

$$\begin{array}{ccc} \text{MacDuff and Richmond, } \underline{\text{the}} \text{ king to be, overcome } \underline{\text{the}} \text{ usurpers and} \\ A & B \\ \text{succeed on } \underline{\text{the}} \text{ political level in restoring } \underline{\text{the}} \text{ order.} \\ C & D \\ \end{array}$$

Since each of the four positions are filled by a "the," a candidate can juggle with the possibilities more readily. The candidate can cover up each example of the article with a pencil or finger. The computer screen is even better: the test can be programmed in such a way that bringing the cursor to the word makes it disappear for the moment. This method requires less imagination of the candidate and probably allows a more concentrated attention on the problem to be solved, raising the validity of such items. After all, when we edit our papers, we also insert words into the text tentatively, just to try them out for size, as it were. We go back and cross them out or erase or delete them when we run through the text once more and see that they failed to work as we had hoped they would.

In a sense, this kind of omi is the opposite of our insertion items, which are also useful in work with advanced learners:

3. Where is the insertion of the definite article *obligatory?*

Candidates are likely to be tempted by the distractors B and C, presumably because a number of European languages require the definite article where school subjects are involved. The standard forms of item analysis, which tabulates the number of times a candidate chooses each of the distractors rather than the correct answer, allows a systematic investigation of what goes wrong

when the item is not answered correctly. Thus it helps us in the class-room to build a bridge from the test item to the principles behind the problem that it poses.

It should be clear that this item format offers a welcome change from standard formats, especially in large test batteries: the change alleviates the monotony of the standard item types. It may also have particular advantages when it is presented on the computer screen. The insertion item on paper requires the candidate to *imagine* the inserted word at four points in the sentence, that is, to reimagine the sentence in four versions so as to come to a decision. The computer, by contrast, can readily present the four alternatives at the touch of a key. Of course this makes the item "cleaner," that is, it obviates the mental operation of imagining the text in a form other than it is and thus facilitates a direct grappling with the problem posed. On the other hand, editing in real life does to a large extent require us to imagine how the text would look after an addition or a deletion, so that the simpler format may be the most valid after all.

One reason why the omi is a good alternative type of item is that it need not be bound to the native language of the candidate: it can be used both with native and with foreign speakers. Consider, for instance, the following item:

4. Find the only word or set of words which can be omitted:

The Emir, $\frac{\text{who}}{A}$ lives in that 450-year-old palace $\frac{\text{which is}}{B}$ painted $\frac{\text{simply on the outside}}{A}$ and furnished $\frac{\text{lavishly inside}}{D}$, drives a $\frac{D}{D}$ babyblue Rolls Royce.

Again, it is only by looking at the sentence as a whole that the candidate will see that the correct answer is B ("which is"). If other changes than those allowed by the four-option format were allowed, any of the distractors would also be potentially expendable. If, for instance, the sentence read "and drives..." then option A would not only be unnecessary but downright wrong; option C is also potentially expendable, except that its excision would cut the ground out from under the contrast inside/outside and so leave option D hanging in mid-air. We have here an item which clearly tests editing skill, then, and one which can be put in a subtest of reading comprehension anywhere, be it for native speakers or for students of French, German or Spanish.

Thus far we have shown items which ask what elements in a sentence can be omitted. What we in fact have more experience with is a rather more sophisticated type of omission item. The head of the item will show why: it asks the candidate to find not what can be omitted but what is *essential* to the sentence, and thus can*not* be removed:

5. Decide which of the underlined word or set of words is essential to the structure of the sentence and could *not* be removed.

No-one $\underset{A}{\underline{among\ us}}$ will ever know $\underset{B}{\underline{the\ reason}}$ why the revolutionary $\underset{C}{\underline{among\ us}}$ movement collapsed $\underset{C}{\underline{just}}$ when its support was $\underset{D}{\underline{widespread}}$ throughout the country.

This item was first tried out in 1979 with a group of university entrants (n = 140) in Cologne. The facility index of .49 seemed satisfactory, and the point-biserial discrimination of over 0.5 most welcome. All the distractors worked at the 5% level. A disadvantage was that both distractors A and C had positive loadings: a number of the candidates who did especially well in the subtest chose these distractors rather than the correct answer. This disadvantage disappeared when the item was included in nation-wide tests for scholarship candidates in the decade after 1982.

Apparently a cluster of six or seven omis in a subtest of reading comprehension represents a factor that is rather different from the conventional "interpret-the-text" item, as the range of discrimination indices will indicate:

TABLE A: Facility and Discrimination

year	n =	fac.	disc.
1982	102	.539	.347
1985	290	.705	.515
1987	229	.867	.422
1989	279	.812	.495
	819		

The instability in the discrimination index is apparently the result of the use of quite different item clusters in the same subtest from one year to an-

other. It is noteworthy that the level of the applicants rose over the decades, thus rendering the item too easy for further use: the increased facility is unrelated to the measuring instrument of the kind of population. The test results themselves, in other words, show why more sophisticated methods have come to be needed: the higher facility index makes the item no longer usable in a test battery meant for university entry.

Now a further turn of the screw is in order, namely the performance of the omission item when not one but one *or* two correct answers are allowed. This is a format which we have also been using for about twenty years, and which is perfectly easy to administer when one uses machine-readable answer-sheets. At first the usual correction for guessing gave us some trouble, but that turned out to be easy to solve satisfactorily. The following is an example of the item type:

Decide which of the underlined words or set of words is essential to the structure of the sentence and could *not* be removed.

6. As to whether the sonnets of Wordsworth are of a character as
$$A$$
 original as those of Keats, the experts disagree. C

Here the candidate is expected to recognize that both A and C are essential to the sentence. In practice, at least 10% of the candidates in the entrance-exam group of 1975 (n= 592) marked either A or C. Again, the item in the Cologne entrance examination was slightly more difficult with a facility index of .285, but it showed a most gratifying discrimination index of .524. It is odd that a quarter of the entrance-exam group were taken in by option B, failing to see that this was deadwood and of no functional use in the sentence.

Fourteen years later the item was included in a scholarship examination for rather more advanced and select students (n= 279). Here over 80% apparently realized that the "As to whether" is needed: at first sight this locution may look like a ready candidate for deletion; but without it the final three words of the sentence make no sense syntactically. So the item turned out to be too easy for this group. The discrimination, however, remained almost exactly what it had been earlier on, namely 0.526, even though the facility index ranged from 0.285 for the group of beginners to 0.826 for end-of-the-second-year students.

This is a gap in facility indices, incidentally, that our reading comprehension items of the more conventional type do not reveal. An investigation of the reason for this phenomenon must be left to a later study. A likely hypothesis is that the more conventional sub-categories of reading comprehension are taught in schools as well as universities. University students, then, may have reached a learning plateau which does not apply to the omi, for that enters the ambit of a kind of editorial competence of which only the more advanced university students gradually become aware. One can attempt to be more precise about what kind of competence this is: it is an acknowledged fact that a beginner revises a paper by looking at local problems; the spelling of a word, the choice of one that is more appropriate or exact, the placement of a comma, the cutting of an overly long sentence into two. The expert edits with paragraphs and suprasegmental structures of argument in mind, considering rhetorical strategies which develop not simply from one sentence to the next but over a multi-paragraph section of the paper. How this is learnt, and whether it is in some direct way teachable, we do not yet know.

One advantage of the omission item in its one-or-two correct answer form is that many of our candidates hope to be teachers or translators or editors one day, and of course no text to be edited ever sends out signals to the effect that each sentence has only a single error or a single correct wording. The idea that there can be more than one error raises the face validity of an editing item without reducing the unusually high discrimination indices. These indices are in turn reflected in high reliability values.

One is tempted to analyse such items in greater detail to see what they actually test. It is a question that factor analysis has not yet made clear, although it shows that the omission item differs from the rest. A nationwide scholarship examination offers a suitable framework for such an analysis because a variety of other domains and item types are included in a test of 230 items and the following item parcels:

- 1. Vocabulary
- 2. Idiom
- 3. Style
- 4. Literary and Linguistic Terminology
- 5. British Civilization
- 6. Literary History
- 7. Reading Comprehension, including omission items
- 8. Grammar

A factor analysis of test results obtained by 279 candidates showed that the cluster of omission items was the only one with a high loading on one factor, whereas all the other parts were more or less dominated by the other factor 1. Subtests 1 and 2 are related to 3, all of the items being in some sense lexical ones, whereas 4, 5 and 6 all test knowledge rather than skills—so it is understandable that as a group they stand somewhat apart from the others in a computerised factor analysis. We must note, however, that factor analysis does not say what the factor is—that is a matter on which we can make an informed guess, based on an analysis of what the items seem to be testing. That it tests editing skills, then, is our thesis, but it cannot be said to have been proved conclusively.

The omission item, then, seems to allow us to test something that is a little different from other standard formats, though it can also be reduced to a set of mere grammar or idiom items. If this is so, proficiency in a cluster of such items seems to be related to vocabulary skills as well. What is probably special about the omi at its best is that it forces the candidate to look at the interrelations between sentence parts. Thus it involves a higher-order skill which more advanced students have to a marked degree, and which characterizes the kind of competence which has helped these students survive a set of hurdles over which some 75% of our students fail to leap. Unfortunately, a study of the prognostic value of the omission items, desirable though it might be, is not possible given the restrictions on time and budget imposed upon us at present.a

WORKS CITED

BAUER, H. 1991. "Sore Finger Items in Multiple-Choice Tests." System 19.4 (1991): 453-458.

BONHEIM, H. 1981. "The Style Test." In *Practice and Problems in Language Testing*. Ed. T. Culhane, C. Klein-Braley and D. K. Stevenson. University of Essex Occasional Papers. 1-17.

BONHEIM, H. and B. KREIFELTS. 1979. Ein Universitätseingangstest für Neuphilologen. Bonn: Bundesminister für Wissenschaft und Forschung.