MULTIPLE CHOICE—TESTING OR SOMETHING MORE?

Mieczyslaw Mendychowski, Politechnika Wroclawska
Wroclaw, Poland

ABSTRACT

KET is a multiple choice software and was developed by three researchers of the Technical University of Wroclaw. It is composed of three independent parts: word processor, test, and results. The first part enables the teacher to prepare 1 question per screen. The question may take up to 250 letters. There may be 4 answers to choose from, of which 1-3 may be correct responses. The second part is a routine test with the possibility of moving the difficult questions to appear later in the queue. The third part summarizes the results both individually and for the whole group. The teacher can immediately correct wrong choices, whether it is a mistake made in preparing the test, or student error. The suggestions for the teacher resulting from this material help optimize the teaching and learning process.

SOME REMARKS ON TESTING AND MULTIPLE CHOICE

Multiple Choice raises an amount of criticism, but as H. Komorowska (1984) remarked: “the type of testing approved is a function of prevailing trends of theoretical views of linguists.” The main objection to Multiple Choice is that it does not involve an active production. This objection is justified. The purpose of Multiple Choice is to test the learner’s understanding skill not that of production. It may nevertheless be observed that Multiple Choice allows an indirect testing of some of the aspects of production skills which is impossible to do in another way: it obliges the student to make a precise choice—a difficulty which he may avoid, consciously or unconsciously, with an open question giving a different answer but an equally satisfactory one.
There is a disparity between the target of learning a foreign language which is to teach to the student a new competence and the impossibility to directly test this competence in another way other than by the means of the performances. Chomsky (Mothe, 1978) challenges the principle of testing itself, but teachers face the problem of evaluation, for administrative reasons (this was not the main Chomskyan worry). There is not a theory of testing but, to my knowledge, there is no theory of language teaching and still languages are taught.

We may wonder how it comes that other evaluations are made and they are close to those obtained by Multiple Choice because indeed there is a high correlation between the said test (Komorowska, 1984) and tests with open answers. It is obvious that a written test allows the whole group to undergo it, (we saved time, though it should be spent on another occasion) we keep copies and we can evaluate objectively much more by computer because it offers speed, memory, accuracy, and patience.

A reproach is made to Multiple Choice as being too easy and giving better scores than open answers for the same contents. It may be due not to facility but to chance and guess. It seems though that the passive knowledge tested with Multiple Choice is more extended than in a case of open answers (ibid, p. 49). One is able to recognize what one can remember the opposite being wrong. Multiple Choice items can even be too difficult, but this feature does not prove the value of items. The questions should be adapted to the level of the students. The answers must not be too long because they become too complicated and the duration of the Multiple Choice should not be excessive and not to go beyond the threshold of tolerance of 90 minutes at most.

In the case of Multiple Choice the student is receptive and is expected to understand the question, and to choose one of the answers prepared by the producer of the test. It is a case of relatively passive recognition since the student does not compose any text. The majority of research on tests evaluating knowledge have shown a very high correlation between the results obtained in a test of comprehension (of Multiple Choice type) and a test of production (the student gives his own answers). Multiple Choice saves time during correction and is more objective, but the open answer question test is easier during its production.
THE PACKAGE KET² AND ITS POSSIBILITIES

The package KET was applied in the following situation. I was a member of a team of teachers who were to prepare a group of adults linguistically to take postgraduate courses in the French-run School of Business. About 50 people had to be evaluated for their knowledge of French and assigned to one of three separate classes. At the end of our one year training scheme they were to undergo an examination and to apply for 24 places at the said School of Business. The computer-assisted part of the evaluation consisted of PIVOT, an oral comprehension evaluation and GRAMAVAN, a grammar test.

PIVOT is an original test (Dumont & Dumont, 1990) recorded by female and male voices, each item has four sentences—one right answer and three distracters. The KET package met those requirements. The text was played two times with a break of 3 minutes. Listening conditions should be the same for everybody for reasons easily understood. When we evaluated the understanding of spoken language we had to take into account the significance of memory. For beginners with only short-term memory functions, and for the more advanced—long-term memory organizes the understanding—the students remember more semantic content rather than the form of spoken text.

GRAMAVAN as meant as a selection test applying the mother tongue—the author did not want to hinder the understanding of the question. Some criticize the application of the mother tongue, but the students appreciated it.

A teacher can produce a classroom test only; a standardized test requires a large, long-distance work team and an adequate number of students. The GRAMAVAN exercises are my own creation, verified by native speakers and teachers of French, and by my colleagues who found some contradictory statements I had simply overlooked. In the case of MASTER the test was prepared in France for foreign students applying for entry in French Universities (SCORE, 1990). Such standardized test should have a sufficient discriminating capacity, i.e. easy for the best candidates and difficult for less good ones. The test is valid if its contents are relevant to the objectives.
The KET package is built with three independent parts:

a) The EDITOR enables preparation of 1 question per screen. The question may take up to 250 letters. There may be 4 multiple choice answers to choose from, and 1 to 3 correct answers. One exercise set may contain up to 200 questions, but in practice 100 is enough. It seems user friendly having recourse to known procedures of prompting, though some skill is demanded on the part of the test producer. Any modification of the written text is easy. Special Polish letters e.g. “ś” are provided. Other necessary ASCII are normally available.

We may choose the exercises or items by the ease of their composition (how to phrase a question and which distracters to find—neither is a simple matter). We cannot test everything by Multiple Choice. The questions and instructions in the items must not be a problem for the student.

b) TEST is a routine test with the possibility of moving the difficult questions to queue later on. Before beginning, the student is asked to enter her last name, first name, and the group’s name. The rules of the game are explained. If the learner wants to “guess” the right answer a “penalty” point will be given; even with this provision, it is improbable that the test will end in a negative score. The student may change his or her choice but only before pressing the “ENTER” key. Each time the questions are chosen at random so that each student faces different problems, thus preventing potential peeking at a neighbor’s screen. The teacher decides when the test is to be stopped, though at the beginning the students are informed of how much time is foreseen to complete the test (usually 90 seconds per question). In practice every person tested who did not know how to answer a question rarely tried his luck at random choosing. Each attempted to guess the right answer.

A new objection may be raised: Multiple Choice favors people with the ability to deduce the right answer from among “other things.” This may also be the case in normal comprehension. A native speaker may adopt a hypothesis which the context makes more or less acceptable. Maybe distracters are of help to our students. That is why distracters should be chosen carefully. Perhaps with the exception of beginners—understanding a foreign language does not happen in an all or nothing fashion, which is the case of Multiple Choice (0 or 1). Despite this limitation Multiple Choice is a valuable instrument in measuring the skill of understanding.
c) RESULTS summarized the outcomes both individually and for the whole group. The teachers receive valuable suggestions for modifying their behavior as well as the student’s. The instructor can immediately correct wrong choices: mistakes made when the test was produced and those made by students for various reasons.

It is true that chance may play a role in choosing right answers but it is possible to reduce its importance to reasonable limits. It may be noted that chance rather favors people with some knowledge of the exercise subject. To reduce the influence of chance, i.e. to reduce the probability of getting the right answers to one chance per million, it is sufficient to increase the number of items to 10 with four possible answers (only one being the correct one. See figure 1.). Since a usual classroom test contains 30 or more items the importance of chance is insignificant. Nevertheless the results should be taken with care; for high values e.g. we do not know how many lucky guesses there are. Also, students sometime speak about “intuition.” In the case of KET the students are warned against “guessing” if they are not sure of the answer. They very much dislike the penalty points!

It is essential to evaluate every item and distracter of the test for their discriminative capacity. Items not satisfactorily discriminating should be removed from the test. The same holds for distracters.

THE EFFECTS DUE TO KET APPLICATION

What were the results of the GRAMAVAN test, and the PIVOT dictation exercise? They were different, of course, but only to a small extent. In my opinion it would not be possible to substantiate placement of every person exactly on the scale after only one GRAMAVAN (38 questions) or PIVOT (16 questions) test. Allow me to pose the
question: how do we know that somebody has understood? There are no criteria for understanding. We can prove that somebody did not understand a fragment but we never can be sure that a learner has understood the whole. Nevertheless, the two tests and the dictation exercise allowed us to create three student groups. There were some minor changes in the “lower level groups.” After our first meeting during normal class time, we had helpful suggestions about what to stress during our first contacts with the groups.

After the results of MASTER some students from the “lower group” who were first willing to undergo the test (the purpose had been stated clearly), later complained about it saying the exercises were too difficult and too stressful. To some degree it is understandable—their results proved rather poor when compared with those who succeeded better. (See figure 2.) MASTER (100 questions) resulted in a different placement for individual students, but given the results, remained within the limits of the group, indicating perhaps more individual involvement in the study. An example of this is the case of two adults who had lived in France for several years, and learned the language without a teacher. They have very good command of spoken French, a very rich vocabulary but still had relatively poor test scores.
Some of the students were using a computer for the first time, and with poor results. After they purchased similar equipment they performed better. I do not know whether or not their personal investment in the computer was the specific reason for their improvement. The final placement of the students who worked very hard in almost all cases confirmed the first results obtained by their first GRAMAVAN and PIVOT placement tests.

All the teachers met after the test and decided that if the success for one question in the case of the whole group was inferior to 60% we should revise. The following list was prepared (reproduces here in part):

1. The spelling of Adverbs—19% success
2. The Adjectives: sensé, censé and sensible
3. The Auxiliary Verbs “avoir” “être” in Passé Composé
4. The Verb faire in construction: on lui fait manger.

After the test MASTER we were bombarded with questions, and I was not always able to satisfy the student’s natural curiosity. Even Grévisse, to my consolation, gives ambiguous answers. But we succeeded in arousing the interest of our students!

<table>
<thead>
<tr>
<th>NR</th>
<th>Max. points</th>
<th>Answers</th>
<th>Correct</th>
<th>Distribution of answ.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>100</td>
<td>62</td>
<td>a* 62 **********</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>b 17 ***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>c 12 **</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>d 2</td>
</tr>
</tbody>
</table>

Figure 3. The analysis of answers pertaining to one question for the whole group
CONCLUSIONS

It can be nice to find defects in the software—it makes one feel better when the intelligent beast proves less informed than it is supposed to be. As to KET—it is a test package. It is not possible to change one’s mind once the answer is entered, such is not the case with traditional on-paper Multiple Choice. The students very much appreciate not only the results but also the possibility of verifying their wrong answers with the right ones. The printer does not print such French letters as é, ê which is a serious nuisance. Precaution must be taken when putting together individual results, a lack of care proved fatal.

In my opinion Computer Assisted Multiple Choice brings a new value to this form of testing especially by giving both individual and collective results at once. Even for beginners in CALL, the hardware becomes, after the first touch, a secondary matter. The individual and collective results interest both the students and the teacher, which creates a natural dialogue and brings out needed modifications from both sides. In the case of small groups the gain of time is not persuasive. To propose tests and exercises correlated with the lessons is very important—hence the need to prepare one’s own materials. Storing the exercises on diskette facilitates their modification afterward. “If the control and evaluation of achievements (…) is the most difficult and unthankful element of the teacher’s work, why not share it with the computer which is a very useful means to simplify the teacher’s task” (Komorowska, 1984).

NOTES

1 Multiple Choice is not an invention of CALL era—it suits well the way the computer can work. The answers should be either completely wrong or correct (0 or 1 values); one short answer alone has no meaning—for the evaluation we need several of them to be representative for the whole batch under examination. Multiple Choice is a set of one question followed by four (usually) answers of which one is correct the rest of them being distracters i.e. false answers but still simulating the right ones. Their role is to make the student think before he makes his choice. (See figure 1.)

A real question and distribution of answers:
1. Quand dois-je vous…ce dossier?
   - a. rapporter
   - b. amener
   - c. emmener
   - d. reporter
The correct answer is shown with an *. In the real case 62% chose the right answer. Still there is some ambiguity for the verbs “amener” and “emmener”. It is an indication for the teacher to work the problem in class the next time the group meets.

KET—Computer Test Examination and Multiple Choice Package was developed by three researchers of the Technical University of Wroclaw. It is suitable to test any possible matter liable to be wrapped in words—that is why this software can be applied to foreign language testing also. This authoring package is composed of three independent parts: EDITOR, TEST and RESULTS. The package runs on IBM and compatibles (512 kb RAM) VGA, EGA. Its authors are: Jaroslaw M. Janiszewski, Jacek Gronowski and Tomasz Piotrowski.

REFERENCES


AUTHOR’S BIODATA

Mieczyslaw Mendychowski, Ph.D., is a teacher of French, Centre of Foreign Languages of the Technical University of Wroclaw, Poland.

AUTHOR’S ADDRESS

Foreign Languages Center
Technical University of Wroclaw
Wybrze e Wyspia skiego 27
Wroclaw, Poland