

# The Capacity of the evaluation of linguistic abilities of FLE examinations applied in Turkey by OSYM \*

**Şeref Kara**

Uludag University, Bursa/Turkey  
serefk@uludag.edu.tr

**Melih Karakuzu**

Atatürk University, Erzurum/Turkey  
karakuzu@atauni.edu.tr

**Abstract :** Since the ultimate object of the assessment and the evaluation in foreign language is the communication competence, it remains custom to determine the parameters and the performance criteria of the foreign language users. However, the multiple choice items used in Turkey generally do not allow us to measure different types of learners such as oral communication. It is limited only to the measurement of mental processes such as knowledge comprehension and application. En the contrary, the multiple choice items are still mostly preferred. Within the framework of this research, our aim is to analyze the multiple choice tests prepared and applied in Turkey by the OSYM in various examinations such as the OSS, the KPDS and the UDS.

**Key words:** evaluation, competence, communication, test, validity.

## 1. Introduction

The evaluation means certain number of various things in various contexts. Sometimes, as teachers, we evaluate our classes as spaces of study and as in teaching; other times, we evaluate the tests or written work that our learners produce.

By many theories of language evaluation the institutional character of the evaluation means is taken into account rather than the learners needs. It is well known that evaluation is an extremely complex field where there are various human and teaching variables. The evaluation helps an educational system to provide the data which is needed to renew itself. So what to evaluate: knowledge or competence? Which type of evaluation use: formative/summative, direct/indirect? How to evaluate: questions/responses, free expression/expression, open questionnaire/questionnaire with multiple choices? It should be noted that it is very difficult to be able to find a real answer miracle, because each teaching method privileges such or such type of evaluation with a precise aim. However, it is not impossible to set up referents to recognize the learners' acquisitions according to equivalences of levels.

The convergent validity with existing tests should not be significant. According to Morrow (1979), it is significant to consider 1) the validity of the contents (items of which the test made up constitute a representative sample for a linguistic capacity?), 2) validity of the psycholinguistic concepts (does the test reflect with precision the principles of a valid theory of a foreign language learners?) and 3) predictive validity (can one determine by the result of the test foreseeable success in a given discipline). False objectivity will not be determining any more, even if, in certain situations it is advantageous to have formats of possible tests corrected by using a machine. Gary Buck (2001) is among the most known specialists for construction and the evaluation in tests in written comprehension. He quotes Bachman and Palmer (1996) by saying that the most significant characteristic of a test is its utility. According to their definition, this concept includes the following features: validity of psycholinguistic concepts, interactivity, authenticity, reliability, practice and impact. Buck adds the effectiveness to it. To build a test and to evaluate it, it is necessary to start firstly by defining the "pattern" or psycholinguistic concept: aptitudes and the capacities on which must carry the test and what it must measure.

In the curricular area of the languages, particularly in programs of FLE, there are many proposals on the way of applying evaluations to the activities of class and the evaluation of program.

The teachers and learners can objectively look at their work and their execution during the course. In the second place, they can understand the progression of the course and its relation with the goals of the program

---

\* The center of selection and placement of the students to the higher education in Turkey

more thoroughly. Thirdly, the teachers and learners can benefit from the evaluation to modify their teaching or study strategies whilst improving their achievements. By taking part in the continuing programme of evaluation, teachers and learners become more implied in the program. In short, the project of evaluation can help to create a good relationship between the teachers and learners.

Continual evaluation is a systematic and reflective process to measure the program. With the difference in other evaluations, this approach comprises only teachers and learners. One can regard it as a process of auto evaluation.

## 2. Scales of levels

In this study, we will present the principal tests, certifications and scales of levels used in foreign language (LT) in Turkey. There are three types of tests in FLT prepared and applied by the OSYM which is the OSS, the KPDS and the UDS.

Intended to choose candidates and historically to place them in the sections of FLE of the universities according to the points obtained and the preferences, oldest of these tests is that of the OSS which is more than 20 years old. The OSS is a test of French language, conceived and created to evaluate the people of an intermediate level. It validates about 250 hours of formation. One can qualify it like a diagnostic test which has a paramount objective to classify and to select the participants. The second, KPDS has existed for 12 years. The third, most recent, the UDS is addressed to those who want to make academic studies after the diploma of licence at the end of 5 years.

The test of the KPDS follows the same format as the test of the OSS and measurement mainly knowledge in general language. It covers subjects accessible to the candidates concerned. It should be noted that the idea of the creation of the KPDS comes from the request of the official institutions for the living languages.

The KPDS and the OSS consist of multiple choice questions which last 3 hours, divided into six sections from 20 to 15 questions each one. The three hours of the test prove indeed often long and discouraging. It is presented in the form of a multiple-choice questionnaire, comprising 100 questions for which only one answer is possible among the 5 choices suggested. The questions are conceived according to a principle of progressive difficulty. But the test of the UDS is composed of 80 questions for which 3 hours also allocated

Though reliable the KPDS is not a reference recognized on an international scale. The total objective is thus well targeted: linguistic abilities in writing.

The certificate given to each candidate indicates a total score spreading out between 0 and 100. But the disadvantage is that none of these three tests provides detailed certificate of linguistic competences in grammar, vocabulary, or written comprehension in order to accompany progress and to personalize the teaching course.

As we announced above the obligatory tests appear as a multiple-choice questionnaire of 100 questions, for a fine evaluation of linguistic competences.

The first part of the tests treats the words of the lexicon: (comprehension of the words according to their usual or rare character); of morphosyntax: (the agreement of the adjectives, the substantives, the past participle, simplest); elements of the nominal group: (determinants, adjectives, possessive phrases), the pronouns in general, relate to the relative ones (differences who/which); the most current times; adverbs of time, place, the prepositions; of syntax: construction negative, interrogative, the simple sentence.

In 2003, there were 100 questions with multiple choices as in the preceding tests. The tests, built by the commission of the OSS, were distributed to the schools with the optical cards of answers and then they were sent to the OSYM in Ankara for the final evaluation. The answers of the participants are corrected and recorded on computers and the results of all the participants are announced on the Internet.

Then the test of the KPDS which aims to evaluate candidates the linguistic level in foreign language in an occupational context is intended to the civil servants, to professionals who would like to know their linguistic level, to companies who would like to determine French level of their current or future employees, it is also intended to organizations of formation for the installation of homogeneous groups or continuous or final evaluation. At the end of the test candidate obtains a certificate of level. The texts of questionnaire relate to the social sciences, economic and legal, mathematical sciences and sciences of the matter, life sciences.

These tests are a measuring instrument of general knowledge in French. They make it possible to get a detailed vision of the learner or employees level in writing and operation of the language. But oral competence is not evaluated.

As KPDS is the validation of knowledge, the certificate is valid only for 5 years, knowledge being able to undergo positive or negative changes. In fact tests give little information would enable us to comprehend the candidate's difficulties. The opinions and interpretations must be changed, and their exchange and evaluation are an essential part of the interactive training, comprising the development of the language, the cultural conscience and the increase of learners in general terms of education. Because very often there is not only one correct answer to a question.

While preparing this work we were relied on our experiments as well as well as on testimonies of the participants of various levels at the exit of the examination rooms. Although a large audience is concerned with these tests, unfortunately we did not find a serious study which relates to them. Normally a tool for evaluation in FLE must measure all the linguistic abilities. But unfortunately the tests of languages of the OSYM completely neglect the oral competence which is of primary importance for the linguistic communication. Within the framework of this work our objective is to highlight some axes which could help the decision makers and the inspectors to change their policy and their approach simply by taking account the communicative function of the language. Rather than to analyze and highlight the gaps or the weaknesses of these tests about which we spoke here, let us try to propose some ways for their future improvement.

### **3. Stages in the evaluation**

The teacher and learners should carry out continuous evaluations periodically. First continuous evaluation can be carried out in the first week.

In a language program, there are many factors which must be considered and studied in order to obtain an objective sight of the process of teaching (Braskamp and others, 1984). These factors can be classified in four categories.

### **4. Gathering of the data**

Manners of gathering data are changed. In our own practice, we employed questionnaires of the tests of the OSYM.

### **5. Feedback and advantages**

The continuous evaluations should not ever be means of sanction for the teacher or learner. They should rather help the two parts in their practices. The teachers and learners them should adopt positive attitudes and concentrate on the achievement of the objectives of the texts, but they can also get information about the models of study and the strategies of learners. In addition, learners find out about teaching style of the teacher and can understand why certain methods are employed. By mutual observations and the exchange of the ideas learners and teachers help each other to improve the learning. Learners can also profit from different studying strategies and the teachers will be able to see learner models and can adjust their methods to satisfy learner needs.

A continuous evaluation is a systematic examination of a language program. By changing the roles of teacher and learner in a program, teaching and the study can be improved. The process should include teachers and learners and various kinds of activities at the various stages of the evaluation. The evaluation also benefits the teacher who becomes a researcher.

All the methods of evaluation have some disadvantages, and it is essential to determine which MCQ can play an essential role in the strategy of total evaluation by examining the results of the study envisaged by module.

However it should be remembered that any form of valid evaluation is likely to require learner to show a certain form of the basic knowledge which is acquired by memorisation. The questions based on basic knowledge are also a formative evaluation tool to check if learner has comprehended it. The preparation of MCQ which tests the knowledge in depth of the learner is more difficult to realize than traditional questions.

### **6. Comprehension and the application**

It is necessary to formulate as clear and concise questions as possible, while avoiding the complex language and the composition of the axes which give more than one indication. It is better also to concentrate on common errors of learner like sectors for questions and/or options. One must accept both answers as correct if two of the choices are possible. But if necessary, if the question requires it, it should be specified that there is only one correct answer.

How can one provide effective feedback for the MCQ? The assistance of feedback can be desirable in the summative and essential evaluation in the formative evaluation. Contrary to the traditional evaluations where feedback depends on various answers, the closed range of response for MCQ means that the tutors know the possible errors which can be made before the test. Thus the proposal of a MCQ is an excellent chance for the tutor to get focused generic feedback. This can be in the form of oral test of the answers following a written test or feedback on the questions, where learners can identify their errors by themselves. The effective feedback does not indicate simply to learners where they were mistaken but also gives the reasons as to why.

## 7. Useful indicators for feedback:

In the best of the cases, it is better to write feedback while writing the questions. The principal strengths and the weaknesses of the multiple choice questions are as follows: they examine a wide range of subjects in little time. The evaluation is not affected by a capacity of writing or orthography of learner. They can be surely pointed while all the answers are predetermined. They can also quickly be marked by computer. The inscription on computer gives the easy access to an analysis of article of the questions in the specific problematic fields for learners. A large bank of the questions which reduces the future preparation time can be made up. They can be employed for the fast revision at the beginning or the end of a class and to be marked by learners.

## 8. Conclusion

One of the advantages of MCQ is the documentation of the statistics on the tests which are easily accessible, especially if these tests are marked by a computer. Multiple choices questions can be employed to examine the comprehension of a reading passage. Inscription of the MCQ, as mentioned above, is often difficult. As for true/false items, they should not directly quote the words of the passage and they should reflect a certain possible ambiguity of the text. None of the possible answers should stand out for example in comparison, having one longer or shorter than the others. None of the possible answers should be the opposite of correct answers. (as that states that usually one of the opposites is the correct answers). The distracters should not have similar meaning, since they can not both be correct, they must both be false.

On the other hand the items should equally examine the information which can be implied from reading passage and require the candidates to reassemble the information from more than one place in the passage. The pre-tests are always significant but in particular with MCQ because obtaining a fresh prospect about the test is significant.

The effectiveness is a feature added by Buck, which stresses the importance to be able to draw from a test as much information as possible on the capacities of written comprehension of the participants within the limited time.

Thus it could be possible to avoid it by making easier tests. In MCQ, it is very significant to write good correct and incorrect replacement choices. The correct choice in question should not come directly from the text. It should be reformulated, so that participant can understand the significance of the text and not simply to identify the repeated words. Incorrect replacement choices should be based on a possible misunderstanding of the text. They should be clearly incorrect, but not illogical. If they are illogical, tested will eliminate them, even if they do not understand the text. The problem in writing goods items is often that it is difficult to write three or four good replacement choices which are logical and clearly incorrect.

By writing MCG, one should avoid giving indications for the correct response the multiple choice test preparers have a tendency to put the correct response to the medium, C-to-D., if there are four choices, they prefer B or C, or if there are five choices C is preferred. One must make sure that the correct answers are not prevalent matter in the medium of the solutions of replacement. As mentioned above, one must also avoid employing the opposites of the correct answers as incorrect alternative, and it is better to employ two incorrect replacement choices with very similar significances.

While preparing the items, one must consider carefully which qualifications or knowledge are necessary to answer the question. If a competence other than that that we want to examine is necessary, this type of item is probably not a good item.

By considering what we said about our test, we can conclude that the test functions relatively well in its context. Principal criticism comes owing to the fact that the written text does not represent the authentic spoken language clearly enough and that the texts could be varied with regard to the topic and the type of text. But one also needs radical changes so that these tests have conformity and an international validity

Finally, there are two significant points that we want to draw from this discussion. First of all, considering the established concept in the specifications of the test, MCT can be useful at least as a part of written comprehension tests. It can show the validity even according to communicative approaches. What is significant and what constitutes a challenge for the specialists. From which our second point comes, is to be able to build a valid test which measures the targeted concept. The difficulty lies in other factors; finding a representative sample of texts and tasks, having as result a degree of interactivity and satisfactory effectiveness. Especially by building a big relatively significant test like the test of the OSS for about of 45000 participants, it is necessary to analyze and check the results before and after the test with people similar to the target group of the test and with the specialists of the discipline, considering the answers and the results obtained.

Conceived and developed by the OSYM, the tree tests of evaluation of French language in Turkey allow testing only competences in written comprehension and completely neglecting competences in oral and written expression in non specialized French language. This causes a great weakness with regard to the reliability and the validity of these tests.

"How can I evaluate myself?", "I am not a teacher!" The traditional school culture does not encourage the learners taking responsibility in the evaluation, the only person who really knows if you understood something is you! At school and university you can often pass from the examinations if you know only 50% of a subject. If you carry out your own test, you can discover what you really know. The evaluation is the responsibility of the teacher at the school and the university partly "yes". However, after the university you'll have to learn from new things for your profession. You'll have to be evaluated yourself to discover what you do not know, so that you can project what you have to learn. That is called "needs analysis". For what is known as in the evaluation of oneself, one needs many supports and counselling from the teachers

The examinations are not a goal but they are means. They measure simply the degree of what learners know. If education has ten essential objectives the examinations and their result constitute simply one of them. But in Turkish education system, the examinations became the only objective in itself. The other objectives lost all their importance. All the actors of education concentrated only on the examinations. The children grow up simply with examinations. They are completely isolated from everyday life and they lost their creativity because of the memorisation system. Everything is evaluated and relied on the result obtained from one examination. The state as well as the parents does not see anything beyond the examinations. The results are not questioned. On the other hand the results are worse for the participants as well as the educational establishments. We insist on something wrong just because of the increase in number of students at the university gates. Because of the examinations the life of the parents is upside down as that of the thousands of youngsters.

Here is a debate for the awakening of the utility of the evaluation and the validity of the written comprehension test. It is clear that there is still a lot to do make in this field, to arrive at the interesting but complex process of written comprehension, and at its valid measurement.

This is why the tests should rather be conceived so as to encourage the recourse to activities of acquisition. The tests of use must measure before all the competence of communication of the learners. For this reason they are especially the abilities of comprehension and communication of the ideas which must be tested.

Contest of selection and placement of the students.

1 Examination of competence of foreign language of the employees of state.

1 Examination of foreign language of the council interuniversitaire.

## References:

- Bachman, L., Palmer, A. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Buck, G. (2001). *Assessing Listening*. Cambridge: Cambridge University Press.
- Bolton, S.(1991). *Évaluation de la compétence communicative en langue étrangère*. Paris: Les Éditions Didier.
- Cadre européen commun de référence pour les langues:apprendre, enseigner, évaluer* (2001) Didier.
- Chapelle, C.A.(1998). *Construct definition and validity inquiry in SLA research*. In: Bachman L.F, A.D. Cohen: Interfaces Between Second Language Acquisition and Language.
- Cornaire, C. (1998). *La compréhension écrite*. CLE International.
- Galissou, R,D.Coste (1986). *Dictionnaire de didactiques des langues* Hachette
- Germain,C.(1993). *Evolution de l'enseignement des langues: 5000 ans d'histoire* Cle Int.
- Hilton,S C.Veltcheff (2003). *L'évaluation en FLE* Hachette.
- Hymes D., (1984). *Vers la compétence de communication*, CREDIF-Hatier.
- Hutchinson, J.A. et Pauline M. Rankin (1987). "Employment profiles and compensation for educational technologists:1983-86" Educational Media and technologie yearbook:1987. Littleton, Co: Librairies Unlimited.
- Landsheere, G. (1992). *Dictionnaire de l'évaluation et de la recherche en éducation*, PUF
- Lhote, E. (1995). *Enseigner l'écrit en interaction*. Paris: Hachette.
- Lussier, D.(1992). *Évaluer les apprentissages dans une approche communicative*. Paris, Hachette.
- Nimier, J. (1996). *La formation psychologique des enseignants*: Collection Formation Permanente en Sciences Humaines. Edition: E.S.F
- Morrow, K. (1979). *Communicative Language testing: revolution or evolution?* in: Brumfit.
- Porcher, L. (1995). *Le français langue étrangère* Hachette.