Investigating the Validity of Achievement Tests in the Algerian Secondary Schools: The Case of EFL Classes in the Biskra Region

Hoadjli Ahmed Chaouki
University of Biskra

Abstract:

This article sets out to explore the issue of evaluation in general and testing in particular within the Algerian context. The motives of such an investigation stems from the students' low scores that do not actually reflect their real performance by virtue of the non-pertinence between what students learn and what they are tested in. Therefore, the main purpose of the present article is to provide a comprehensive evaluation of the achievement tests to the EFL classes in the Biskra region.

الملخّـــــص:

يدور مضمون هذا المقال حول موضوع التقويم التربوي بصفة عامة و إشكالية بناء الاختبارات الفصلية بصفة خاصة بالمدرسة الجزائرية و تحديدا بالنسبة لطلبة الثانوي بولاية بسكرة ، تطرقنا لهذا الموضوع بالذات نابع من ملاحظاتنا الدائمة حول عدم وجود تطابق بين محتوى الاختبارات المقدمة لتلاميذ هذا الطور و ما يدرس خلال الفصل أو السنة الدراسية .و عليه فإن الهدف الأساسي لهذا المقال هو إجراء عملية تقويم لهذا النوع من الاختبارات و الوقوف على طريقة بنائها و ذلك من أجل تحديد النقائص و اقتراح بعض الإصلاحات لجعل هذه الاختبارات ذات فائدة.

 Introduction

Language tests are regarded now to be a valuable tool for providing information that is relevant to several concerns in language teaching. They can be one way of providing systematic feedback for both teachers and students. The teacher can see how well or badly the students are performing and check for any discrepancies between expectations and actual performance. Likewise, the students can know how much attainment and progress they are doing in learning the language.

Tests can also be a good means in evaluating instructional materials and tasks and their relevance to the objectives set out at the beginning. Ideally, the goals of tests or test items should be clear to students, so that they need not spend guessing what the teacher means. If students perceive the test as relevant to their needs in the course, they themselves are probably going to engage more actively in the process of dealing with it.

Another aspect of language tests concerns the insights and inferences a teacher often draws from the outcomes of tests. The usefulness of such inferences is manifested when they provide feedback to be utilized in making the teaching programme more effective and when they provide information as to what kinds of particular materials and tasks students need. In a nutshell, these inferences can be the only ground on which teachers can make appropriate decisions about the teaching operation.

But still, to accomplish all these tasks, a language test must meet the requirements of some basic qualities such as validity, reliability, and practicality. In other contexts, it should be authentic, interactive, and should have impact on all the concerned participants and related areas. Certainly, these considerations will vary from one specific situation to another because what might be appropriate in a given case might not be so in another. That is why an understanding of how test developers come to design and develop a language test is really crucial.

1. Theoretical background

Some important testing concepts are used throughout this article. For a layman, all these concepts may have the same use. And for some of them, they can be used interchangeably. But, for someone whose interest is educational practice or research, these particular concepts bear quite different senses; so, we usually need to put a clear-cut

definition to each concept in order to avoid any misuse or ambiguity and hence not to be misled. Most of all, it is important to understand the logic for the use of each concept to facilitate access to actual applications of each concept in its specific appropriate context.

1.1. Evaluation

It has always been argued that the concept evaluation is not limited to instructional educational contexts, but also transgresses the contexts to be applied in our daily lives. In its broadest sense, evaluation is a natural activity that can be formal or informal. It is something that may not always be made explicit, but may actually be undertaken unconsciously (Rea-Dickins & Germaine, 1992: 02). It is not always something we do in a principled and systematic way. The criteria we use in making judgment may sometimes be vague and ill- prepared. In education, however, evaluation has different aspects. It becomes important to make explicit the criteria used in our judgment. Ill-prepared and inconsistent evaluations are unfair, uninformative, and unreliable (ibid). Evaluation is a process through which teachers judge the quality of their work, their own, or their students. Thus, it is important for the teacher because it enables him/her to obtain a wide array of information to apply and use for classroom practice, in addition to being useful for learners in order to enhance their learning tasks. The principal aim of evaluation, in this context, is to make sound choices. It is then concerned primarily with decision making (Richards, 1996, cited in Cohen, 1980:03). Such decisions are based on informed judgment. They need to be careful in collecting information and thoughtful about interpreting that information (Genesee& Upsher, 1996).

1.2. Testing

Testing, at large, is a universal socio-cultural phenomenon of fundamental importance of test takers because it has a great impact on their lives. What is surprising, given the validity of testing, is that its practice is so little understood (McNamara, 1996). Language testing has the form of negative connotation, such as feeling of failure and lack of self-esteem which may last beyond school-days (Hedge, 1993). Testing is considered as one component in the evaluation process. Hence, if evaluation is concerned with informing decision making, it is difficult to imagine rational decisions to be made without tests (Hughes, 1989). Caroll (1968) makes this last assumption clearer enough, and defines tests as ' a procedure designed to elicit certain

behavior from which one can make inferences about certain characteristics of an individual'(p. 46).

A test, then, is a method in the sense that it includes techniques, procedures, and tasks, which constitute an instrument of some sort, and that method needs an activity on the part of all the participants in the testing operation. More precisely, testing is a method of measuring a person's ability or knowledge in a given area (Brown, 1987). It is also designed to measure the learner's competence in language at a particular moment in the course and nothing else (Corder, 1973). This is supposed to measure in order to formulate a judgment that has to be interpreted. To make this interpretation valuable, it must be quantified in a more or less exact way (Bachman, 1997). This quantification has to be applied and performed according to explicit rules and principled ways because ad-hoc testing is likely to be uninformative and unsuitable source on which to base inferences and to make educational decisions.

1.3. Assessment/ measurement

In the literature on testing, assessment and measurement are used interchangeably. What is true for testing can also be used for the two concepts. If testing is the procedure of measuring the learners' abilities, assessment is the score yielded by this procedure. Assessment, in its precise sense, refers to the process of quantifying the characteristics of a person (ibid). Quantifying, then, means measuring in a mathematically precise terms the ability or area to be tested. Bachman points out three different distinguishing features of assessment: quantification, characteristics, and explicit procedures.

First, quantification is the assigning of numerical or non numerical categories. Very often, the numerical categories are presented in terms of figures, whereas the non-numerical categories are displayed in the form of letters or labels. Both ways of categorization aim at providing accurate analysis and interpretation. Second, characteristics concern mainly mental attributes and physical abilities. Caroll (1968) has proposed defining ability with respect to a particular class of cognitive or mental tasks that an individual is required to perform, and mental ability thus refers to performance on a set of mental tasks (p. 268). The particularity of mental abilities is that they are not directly observable, and hence pose a certain difficulty for test takers to measure them. Such kind of mental abilities include examples such as fluency in speaking, achievement in reading, motivation, intelligence,

and so on. On the other hand, physical abilities do not create any trouble since they are observable and therefore can easily be measured. The third distinguishing characteristic in assessment is the one concerned with quantifications and how they should be performed. Generally, the procedures and techniques to be followed have to be clear and well-structured. In other words, this simply means that when quantifying there is no room for arbitrary or blind assigning of grades to the characteristics of individuals; rather, these should be based on principled criteria.

1.4. Achievement Tests

This type of tests is used to measure the extent of learning in a problem described domain often in accordance with explicit stated objectives of a learning programme (Henning, 1978). The purpose of achievement tests are categorized into two main areas (1) - general objectives, and (2) - specific purposes. The general objectives are often an integral part, or go in parallel with the course objectives and instruction. Hughes (1989) insists on the idea that the main concern of achievement tests is to establish how successful students, or groups of students, or the courses themselves, have been in achieving objectives. This means, in other words, one needs to see whether there is a mastery or a non mastery of the range of skills taught in a course and whether a progress has been and the extent to which tests are supporting the teaching to which they relate.

Likewise, if one assumes that a well-planned course should measure the extent to which students fulfilled course objectives, then, the specific purposes of achievement tests can provide detailed and thorough information about the degree to which learning and instruction are sustainable and integrated to the goals of the course. Some of these purposes can be identified as determining how well learners are progressing. This can stimulate them to take learning more seriously. In this case, the test tells teachers what can students or cannot do. In other words, the test show teachers how well successful their teaching has been. It provides feedback for them to adjust and change course content and teaching styles where necessary. By identifying weaknesses and strengths, achievement tests can help identify areas of remedial work. Moreover, this test permits to have an accurate judgment of each student in relation to the progress he/ she is making through comparison of his/ her score with the other learners' scores. Also, it helps to evaluate the effectiveness of the programmes, course books, materials and methods used by the teacher

(Carmen, 1995). And more importantly, meaningful feedback drawn from achievement tests can actually yield a solid background for a set of decisions that can help to enhance instructional practice and students' learning with respect to the goals. Such kind of decisions often concern directly students and instruction. Decisions about students include progress and grading to which students meet minimum standards of mastery of the content of a simple instructional unit. Decisions about instruction include what proportions of unit have been taught and the proportions that might require review (Bachman& Palmer, 2000).

1.5. Test qualities

The primary purpose of a useful test is to provide an accurate measurement of the learners' language abilities or how progress learners are doing in achieving language knowledge. In order to prove its usefulness, a language test needs to meet the requirements of some qualities. Myriads of such qualities have been suggested by language test developers to evaluate the degree of usefulness of any particular test. For instance, Lado (1959) states that a valuable test is the one which is said to be valid, reliable, scorable, economical, and administrable. Weir (1988) emphasizes on that concepts such as validity, reliability and efficiency can strongly affect all aspects of test design. Recently, Bachman and Palmer (2000) have provided a very remarkable framework in which they have argued that the most important considerations in designing a useful test can be defined in terms of six qualities: validity, reliability, practicality, interactiveness, authenticity and impact.

1.5.1. Validity

The concept validity is often attributed to the degree to which a test actually measures what it is supposed to measure (Hughes, 1989). It is the test procedure that concerns the extent to which it does what it is intended to do (Philliner, 1968). It is the truth of the test in relation to what to evaluate. It concerns the relevance and usefulness of what one is measuring. The aim of validity is to provide consistent measures of precisely the abilities one is interested in.

1.5.2. Reliability

Reliability is about precise and repeatable measurement on a clear basis of this measurement. It is the consistency of the test's judgment and results (Good& Brophy, 1991). This means that the test should be dependable in its content as well as in its scores. In other

words, if given to the same group of students on different occasions, the test should provide similar results, and if corrected by two different teachers, it should give dependable scoring; that is, the more similar scores are, the more reliable the test is.

1.5.3. Authenticity

Authenticity is defined as the degree of congruence between the test situation and the real-life situation that the student is supposed to master according to the curriculum. In this way, the major concern of this concept is the relationship between the test content and test tasks with the content of what students and teachers have performed as learning process and instruction. In other words, this means that it is quite plausible that a student should be tested merely on something that is characteristics of the language situation use, which is part of the curriculum that has been taught with.

1.5.4. Interactiveness

Interactiveness is often determined by considering the characteristics of the students in relation to the tasks characteristics. That is, test tasks' characteristics can have a major role in the outcomes of the test tasks' performance. Likewise, the kind of information present in the test tasks can have its effect on how test takers perform on a language test (Riley& Lee, 1986). Generally, test takers' characteristics concern characteristics such as academic, background, culture, gender, and field dependence (Kunnan, 1988); other characteristics that need to be taken into account even with a relative interest include aptitude (Sasaki, 1986; Sparks et al., 1998), background knowledge (Calapham, 1993), and personality characteristics (Berry, 1983). Along with the interest of the students' characteristics on their performance, test takers characteristics, too, have to affect how students perceive test content and which strategies students employ to respond to the different test tasks. These test tasks often refer to examples such as the negativeness, concreteness, topic specificity of information, level of vocabulary, syntactic complexity, and cognitive demand, or amount of processing required (Lew-Kowicz, 2000).

1.5.5. Impact

The impact of a test is not restricted to the test itself alone, but it can also affect other areas that are closely or loosely related to the test. Bachman (1997) refers to the impact of a test at two levels:(1)- a micro-level, in terms of the individuals who are affected by the particular test use; and (2)- a macro- level, which concerns a broader

level that is related to the educational system and even transgresses it to touch society at large.

1.5.6. Practicality

Unlike the previous test qualities, practicality is concerned particularly with the ways to implement a test. A test which is said to be practical is a test which requires some necessary resources. Generally, these resources are classified into three categories. First, human resources which comprise the teachers and test administration. Second, material resources often presented in terms of space, equipment and materials. Finally, the third category concerns the time allocated for the test from the initial stage in the process of designing till the last stage in the whole operation.

2. The study

2.1. Research purpose and questions

The general purpose of this study is not to criticize the existing achievement tests in the Algerian secondary schools, for the sake of criticizing them, but the researcher believes that this investigation will be a good opportunity to provide an objective evaluation of how secondary school English teachers in the Biskra region design and develop their achievement tests in order to judge the degree of attainment and progress their students are making in accordance with the contents of the syllabus and their subsequent instructional objectives. Therefore, on the basis of this general purpose, the study will have an ultimate goal setting satisfactorily answers to the following questions.

-       Main Study Question:
- What is the degree of congruence between the typology of instructional tasks and the content of achievement tests used in the Algerian EFL classes?
-       Sub-Questions
- What are the skills covered by achievement tests?
- What is the degree of reliability and validity of these tests?
- How do teachers balance test development and classroom characteristics?
- Do these tests take into consideration the cognitive abilities and the psychological makes-up of students?
- To what extent do these tests reach the final objectives set out by teachers?

- Do teachers in the Algerian secondary schools identify as precisely as possible when developing achievements tests the following matters:
    - The purpose of the test?
    - The characteristics of students?
    - The constructs to be measured?
    - The methods of testing?
- Do teachers provide test specifications when developing their tests?
- Do teachers follow the different procedural stages when developing their tests?

2.2. The Method

Since this investigation sought to provide an account for the basics on which EFL teachers in the Biskra region come to design and develop achievement tests for their classes, the descriptive method is regarded to be the most appropriate. In its essence, this method allows the researcher to opt for some particular strategies and procedures that enable him to remain objective and positive and avoid any subjectivity or bias throughout the different stages that this study took. It is believed that the shortcomings of this method are few and its adoption is easy and time consuming.

2.2.1. The Participants

The main participants in this study were EFL secondary school teachers in the Biskra region. What is worth noting about this matter is that carrying out an investigation on all the secondary school teachers in this region would be difficult to accomplish since this requires time and effort on the part of the researcher. In addition, the nature of this study, in that, it is a case study, does not obligatorily need to pay attention to the whole population since it is argued, as far as research methodology is concerned, that even if a case study yields researchers with a description of the situation to capture the full complexity and uniqueness of information (Ladico, Spaulding&Voegttle, 2007), its merit is that it does not seek for a generalization of the obtained results.

Therefore, on the ground, because the sample was limited to the EFL teachers in the Biskra region, the researcher opted to select (07) teachers from a population that comprised (15). The choice of these teachers was according to a sampling which consisted in choosing the teachers on the basis of their experience in teaching. It is believed that in the case of test design and development experience plays an important role. Of course, this does not mean that the so-called 'green'

or ' novice' teachers could not stand as reliable informants. Rather, the selection of the most experienced teachers is interpreted by the assumption that it is easier with them to gain a wealth of information that can likely help the researcher in answering the myriads of the raised questions in the study.

2.2.2. The Instruments

Since in the present study the focal point to deal with is to provide a description of how EFL teachers proceed to design and develop achievement tests for their classes, the researcher thought that the appropriate data collection instrument for this study is to collect information through a semi-structured questionnaire for teachers. Indeed, if properly designed, a semi-structured questionnaire may yield us with useful data with a large population. Dörnyei (2003) defines this type of questionnaires as ' any written instruments that present respondents with a series of questions of statements to which they react either by writing out their answers or selecting from among the existing answers'(p. 06). He also adds' semi-structured questionnaires are especially valuable because they are efficient in terms of (a) - research time, (b) - researchers' effort, and (c) - financial resources'(p.09). Some of the criteria of a good questionnaire are that it should be typed neatly and designed carefully (Bell, 1987). And the objectives of the questionnaire should be clearly stated.

Because the teachers' questionnaire was intended to provide a comprehensive evaluation of achievement tests for EFL classes in the Biskra region, the relevant specific aims were to elucidate:

1) the pertinence between achievement tests and the teaching programme,
2) the way EFL teachers design and develop this type of tests, and
3) how these teachers perceive their tests.

In terms of its structure, the questionnaire was sectioned into three separate parts:

- ✓ Part One: Test Design.
- ✓ Part Two: Test Development.
- ✓ Part Three: Teachers' Perceptions to these Tests.

The questionnaire comprised (36) items. They were arranged as follows:

-      Part One: Test Design.
        TQ1: Testing Concepts.
        TQ2: Purpose (s) of the Test.
        TQ3: Students' Characteristics.
        TQ4+TQ5+TQ5: Constructs.
        TQ6+ TQ7+ TQ8+ TQ9
        TQ10+ TQ11+ TQ12+ TQ13 ⟶ TestQualities

        TQ14+ TQ15+ TQ16
-      Part Two: Test Development.
        TQ17+ TQ18+ TQ19: Test Format.
        TQ 21+ TQ22+ TQ23+ TQ24: Test Specifications.
        TQ25+TQ26+TQ27: Test Techniques.
        TQ28+Q29+TQ 30: Major Skills.
        TQ31+TQ32+TQ33: The Scoring Method.
-      Part Three: Teachers' Perceptions.
        TQ34+TQ35+TQ36: Teachers' Perceptions towards the Current Testing System.

To make the final questionnaire ready for use, the researcher proceeded to try-out a first version of this questionnaire. The piloting of this questionnaire was held through two separate phases. In the first phase, a preliminary questionnaire was designed. It comprised a wide range of items and techniques of questioning. The aim of the first try-out was to select the appropriate items for final use. Besides, it was intended to evaluate the degree of instructions' clarity, the lay-out's attraction, the relevance of the items to the content of study, and if the teachers objected to answering some items. The second questionnaire is the final version.

3. Results

-      When asked about the definition of some concepts common in the testing literature such as a 'test', 'test design', 'test development' and others, the researcher got different responses. What sorts out from these responses is that the great majority of teachers provided personal explanations, and most of these explanations were mere subjective interpretations. Besides, these answers indicated that the teachers could not make clear-cut definitions among the various concepts that are available on the testing literature.
-      When asked about what types of information these tests provide you with, more than the half of the respondents thought that these

tests yielded them with the students' level. The rest of the respondents said their tests usually can help them too assign scores for their students in order to see who are going to pass and the ones who are going to fail. Surprisingly, none of the respondents indicated that achievement tests aid them to make an evaluation of their students' attainment and progress in accordance with the instructional objectives.

-     When asked whether or not they base their tests on any theoretical basis, none of the respondents gave an affirmative answer. All the questioned teachers left this item unanswered.

-     When asked whether or not the scores their students obtain reflect their actual level, only one third of the respondents said that actually the scores reflect the level of their students. The other two thirds were not sure.

-     When asked whether or not they consider the affective side when designing achievement tests, the majority of the respondents answered negatively to this question. They argued that giving too much importance to affection will make this type of tests lose its validity and this will obviously lead to non-credible scores.

-     When asked about the typology of tasks their tests usually comprise, the respondents were unanimous to answer that the majority of tasks turned around questions on grammatical structures.

-     When asked about whether or not they test the major two skills: listening and speaking, all of the respondents agreed on the fact that they never test these skills. Their argument was that they do not test these two skills because they are not recommended in the official 'Examination Guide'. The latter corresponds in its format and contents to the 'BAC' examination model. For them, adopting this model gives their students a good opportunity to be trained on the typology of the tasks that often come in this national examination. Besides, the teachers who were questioned added that some factors such as the overcrowded classes and the shortage of resources prevent them to incorporate these skills in their tests.

-     When asked to display their perceptions on the subject of testing, the great majority of the respondents said that they did not give too much importance to this matter, first, because they conceive the task of testing as an operation where the most important element is to assign scores for their students. Second, these teachers ignore completely the systematic way to develop a test because they did not get any training in their careers about this issue.

-     At the end, when asked about whether or not the current testing system adopted by EFL teachers in the Algerian secondary schools,

should be kept, or modified, or completely changed, most of the respondents came up to the assumption that it is better to review this system and strive for making it more suitable to the demands of the present time and mainly to the recent communicative teaching approaches.

4. Discussion

In answer to the main research question and its relevant sub-questions, it is very conspicuous to say that achievement tests for EFL classes in the Biskra region, as they are constructed by secondary school teachers, do not fit the requirements of a useful test. In particular, these tests do not rely on any theoretical background and do not follow any operational and procedural ways. Hence, teachers completely ignore what to put in the crucial stages in test design and development. In more precise terms, the EFL teachers, question to the present study, do not care about the purpose or purposes of the test, do not identify as clearly as possible the constructs to be measured, and totally neglect the aspect related to the characteristics of students. At another level, the same attitude is expressed. That is, teachers do not pay attention to the relationship between the test format and its content, in that these teachers do not adapt the divisions and sections that make-up an achievement test to the number of test tasks and items that students are asked to answer. Consequently, one can easily pinpoint that there is disequilibrium in the distribution of tests tasks and items among the various parts of a test. One section seems longer than the others. Moreover, when it is up to yield a scoring scale, what stands out clearly is that the scores are not balanced. In terms of test tasks and items difficulty, what is remarkable is that there is no gradation in designing the different questions of the test. Ostensibly, the way of eliciting the test tasks and items do not correspond to Bloom's taxonomy, a fact that renders the issue of answering these questions very difficult on the part of students.

Furthermore, when time comes to the development stages, what is worth to mention on this point is that teachers do not know exactly what the matter of 'test specifications' is. For the great majority of those teachers, they do not refer to this crucial step in test development to make their tests resemble to what they teach in class. Instead, it sounds that to develop their tests, these teachers only pick up ready-made tests from the available commercial books; or they go back to previous constructed tests; or they simply base themselves on their intuition to construct tests. Worse than this, these tests that are

often constructed in a hurry are examples that narrow the scope of teaching and learning since their contents almost comprise quite the same typology of tasks that are found in the ' BAC' examination. On this last idea, teachers would argue that they usually proceed in this way because they are often judged on the results their students obtain in the national examination. On the ground of this, they think it is more pragmatic for them to train their students to answer questions that look like the ones that are found in the 'BAC' exam before time comes to sit for this decisive examination.

Therefore, based on these notes and the above diagnosis, it is worth noting to point out that, at the level of the present investigation, the subject of testing remains something that needs to be treated in the strictest care and consideration. This can solely be achieved by giving a great importance to some specific qualities such as validity, reliability, practicality, authenticity, interactiveness, and impact. What is then recommended is to make some effort to maximize the attainment of a certain balance among those qualities since it is actually very complicated to meet the requirements of each quality individually in all occasions and instances. In this respect, this investigation has provided us with evidence that there is a need of developing tests that correlate with the instructional goals. There is a need of developing tests that better identify the students' specific needs and strengths, so that teachers can be appropriately informed about the students' achievement. There is a need to avoid any sort of influence or authority of external examination on instruction and learning and that often lead to a negative washback. Rather, there is a need to develop tests that must exert a positive impact on both teachers and students and make teaching more effective and learning positive.

5. Recommendations for Pedagogy

As a response to the myriads of anomalies that were pinpointed in the teachers' practices, the researcher has thought that it would be better to devise the following set of pedagogical recommendations for EFL teachers in the Algerian secondary schools about how they should proceed to design and develop useful achievement tests:

- The teacher has to identify as clearly as possible the purposes of the test. Generally, this concerns knowing how students are progressing, finding their weaknesses and strengths, working out ways of helping

them, and using information from the assessment for the planning of their classes.

- Through the above activity, the teacher could be able to explicit the specific uses for which the test is intended. It involves clearly stating the inferences he/ she intends to make on the basis of test results, and any specific decisions he/ she will make should be based upon these inferences.

- The teacher has to rely on clear theoretical bases when designing his/ her tests. The advantage of basing a test on a clear theoretical background will definitely allow teachers to become competent in developing and using achievement tests for the intended purposes.

- Basing a test on clear theoretical principles allows the avoidance of intuitive and ready-made tests that are irrelevant to instruction and which provide inconsistent and inappropriate inferences about the final outcomes.

- The teacher should be creative and flexible, and avoid being monotonous and fixed on the same test each time he/ she is asked to develop one.

- The teacher should develop as clearly as possible the language abilities and skills to be taken into account when designing his/ her test. By defining the language abilities/ skills in this way, the teacher is making clear the definition of the constructs to be measured.

- Another consideration is to decide which specific components of language abilities in syllabus are to be included in the construct definition. In many cases, the teacher wants to make inferences about specific components of language abilities and may thus define the constructs in terms of those components. This might be the case if a test is to measure the degree of progress and achievement of specific syllabus objectives. The teacher will then most likely base the construct definition on specific language abilities that are to be included in the syllabus.

- The teacher has to consider the learning styles of his/ her students. Such a consideration is often determined by the importance allowed to the cognitive/ psychological abilities and makes-up of students. This can be made possible through the design of achievement tests following Bloom's taxonomy.

- The teacher has to take into consideration the affective side of his/ her students when designing and developing tests. This allows the elimination of all sources of frustration and anxiety. A positive attitude

vis-à-vis the test will help achieve better scores and make substantial progress.

- The teacher has to be usually aware of the importance of considering some crucial test qualities such as validity, reliability, practicality, authenticity, interactiveness, and impact. The teacher needs to strive to create some balance among these test qualities.

- The teacher should not neglect the students' characteristics when designing and developing achievement tests. This mainly concerns personality characteristics, background knowledge, aptitude for learning, and some other matters.

- The teacher has to treat the students as responsible individuals by providing them with as complete information as possible about how a test is developed and the scoring procedure followed.

- The teacher has to consider the testing progress as an integral part of the teaching operation. The relationship between teaching and testing should be regarded as that of partnership.

Conclusion

To put it in a nutshell, from the present investigation, we have realized that achievement tests used by English teachers for EFL classes in the secondary schools in the Biskra region are not appropriate to make a comprehensive evaluation. Most of these achievement tests are ready-made tests, or external examinations, or merely tests developed by teachers in a hurry and by intuition. They do not assess the developed skills in instruction. They do not cover a considerable amount of the syllabus-content. They are not developed through systematic stages. And worse than these, they are not related to clear, specific goals and general objectives. In some cases, these tests are useful tools for evaluating a part of the programme, or certifying general competencies, but less helpful in diagnosing students' needs and strengths in different stages.

As a remedy to these deficiencies, the present study has displayed for us that the issue of testing is not an easy matter. Rather, it is a very systematic process and a well-principled procedure of gathering data about general and specific language abilities from performance on tasks designed to provide a basis for consistent and comprehensive interpretations of learners' outcomes. Out of this investigation, we have come to assert that teachers must conceive testing as an integral part of teaching and the relationship between the two must be regarded of partnership. And finally, we have strongly to emphasize

on the dire need to consider testing as a source of collecting useful feedback as how to improve teaching and enhance learning.

References

- Bachman, L. F. (1997). Fundamental consideration in language testing. (4th. Ed). Oxford, New York. Oxford University Press.
- Bachman, L.F., & Palmer, A.S. (2000). Language testing in practice. Oxford, New York. Oxford University Press.
- Bell, J. (1987). Doing your research project: a guide for first time researchers in education and Social sciences. Milton, Philadelphia: Open University Press.
- Berry, V. (1983). Personality characteristics as a potential source of language test bias. In Alderson, J. C., and B., editors, language testing, 17 (1). 1-10.
- Brown, H. D. (1987). Principles of language learning and teaching. Eaglewood cliffs, NJ: Prentice-Hall, Inc.
- Calapham, C. (1993). The Effect of language proficiency on EAP students' reading comprehension. In Alderson, J. C., and B., editors., in language testing, 17(1), 1-10.
- Carmen, P.B. (1995). Coming to grips with progress testing. In forum magazine, 33(3), 09-55.
- Carroll, J. B. (1968). The psychology of language testing. In A. Davies (Ed.). Language testing Symposium: A psycholinguistic approach. London. Oxford University Press, 06-46.
- Cohen, A. D. (1980). Testing language ability in the classroom. Rowley, Mass: News burry house.
- Corder, S. P. (1973). Introducing applied linguistics. Middlesex, England. The Penguin Books Ltd.
- Dörnyei, Z. (2003). Questionnaire in second language research: construction; administration, and Processing. Mahwah; NJ: Lawrence E and L Baum Associates. In TESL-EJ, 6(3).
- Genesee, F., & Upshur, J. A. (1996). Classroom –based evaluation in second language education. Cambridge, U. K: Cambridge University Press.
- Good, T. M &Brophy, J. (1991). Educational psychology: A realistic approach. (4th Ed.). Longman.
- Hedge, T. (1993). Key concepts in ELT fluency. In English teaching journal, vol (47), 275-277
- Henning, G. (1978). Developing English language proficiency measures for native speakers of Arabic. UCLA work-papers in teaching English as a second language, vol (12).

- Hughes, A. (1989). Testing for language teachers. Cambridge, U.K: Cambridge University Press.
- Kunnan, A. J. (1988). Approaches to validation in language assessment. In Alderson, J. C., & B., Editors, language testing, 17(1), 11-2.
- Lado, R. (1959). Language testing. New York: Mc Graw- Hill.
- Lew Kowicz, J. A. (2000). Authenticity in language testing: some outstanding questions. In Alderson,J.C. and B. Editors, testing 17(1), 3-12.
- Sasaki, M. (1986). Second language proficiency, foreign language, aptitude and intelligence: quantitative and qualitative analyses. In Alderson, J. C& B. Editors, language testing, 17 (1), 1-10.
- Sparks, R. L., et al. (1998). Differences: a native language skill, foreign language aptitude among high –average, and low-proficiency language users: two studies. In language testing, 17(1), 01-10.
- Rea- Dickins, P., &Germaine. (1992). Evaluation. Oxford: Oxford University Press.
- Rirely, G. L., &Lee, G. F. (1996). A comparison of recall and summary protocols as measures of second language reading comprehension. In Alderson, J. C. &B. Editors, language testing 17(1), 01-10.
- Weir. J. (1988). Communicative language testing. Exeter linguistics Studies, vol (11). University of Exeter.
- Mc Nammara, T. (1996). Second language performance measuring. London and New York: Longman.
- Philliner, A. G. (1968). Subjective and objective testing. In A. Davies (Ed), language testing symposium. A psycholinguistic perspective. London: Oxford University Press.