

A PLACE FOR EVALUATION, ASSESSMENT AND TESTING IN FOREIGN LANGUAGE TEACHING

Ana Clara Sánchez S y Gabriel Obando G.
Universidad de Nariño, Pasto⁴

ABSTRACT

This article presents an outline of some key concepts in the field of assessment. It also sketches the relationship between assessment and instruction and the implications of designing assessment tasks effectively. The principles and steps in creating these tasks are also addressed as well as the challenges foreign language teachers will have to face in the near future regarding assessment practices.

⁴ Ana Clara Sánchez Solarte es Profesora Asistente del Depto. De Lingüística e Idiomas. Posee una Maestría en enseñanza del Inglés como Lengua extranjera de la Northern Iowa University, una Especialización en Traducción y otra en Docencia Universitaria de la Universidad de Nariño.
Gabriel Obando Guerrero es Profesor Asistente del Depto. De Lingüística e Idiomas . Obtuvo su Maestría en Enseñanza del Inglés como Lengua extranjera de la Northern Iowa University y también posee dos especializaciones, una en Traducción y otra en Docencia Universitaria de la Universidad de Nariño.

KEY WORDS

Evaluation, assessment, testing, foreign language teaching

RESUMEN

Este artículo presenta un de algunos conceptos claves en el campo de la evaluación. Así mismo se esboza la relación entre la evaluación y la instrucción y las implicaciones de diseñar actividades evaluativas efectivas. Los principios y los pasos necesarios para al creación de estas actividades evaluativas también se incluyen en el documento, así como los retos que enfrentarán los docentes en un futuro cercano dadas las nuevas tendencias en evaluación.

PALABRAS CLAVES

Evaluación, actividades evaluativas, enseñanza de lenguas extranjeras

Sometimes teachers seem to forget that teaching is not the center of our professional practice. It is evident that the amount of teaching carried out in classes does not equal the amount of learning taking place (as shown by the divergence in the learners' proficiency inside one same group). It is important then, that foreign language teachers rethink their role in the teaching and learning processes and, most importantly, that they understand the multidimensionality of foreign language learning.

One of the aspects that foreign language (FL) educators need to readdress is the role of assessment in learning. Judging from what can be observed in foreign language classrooms one might say that assessment is seen by educators as a disconnected activity which bears no clear relationship or influence on learning; an event that takes place every once in a while and is as tedious for teachers as it is for students, and which is immediately forgotten after it has passed. In fact, some students may even say that assessment is a frightening experience they want to put out of their mind the minute they step out of the classroom. Assessment seems to fulfill the only purpose of getting some scores to be shown to administrators at the end of a term. This misconception needs to be clarified in order to enhance the

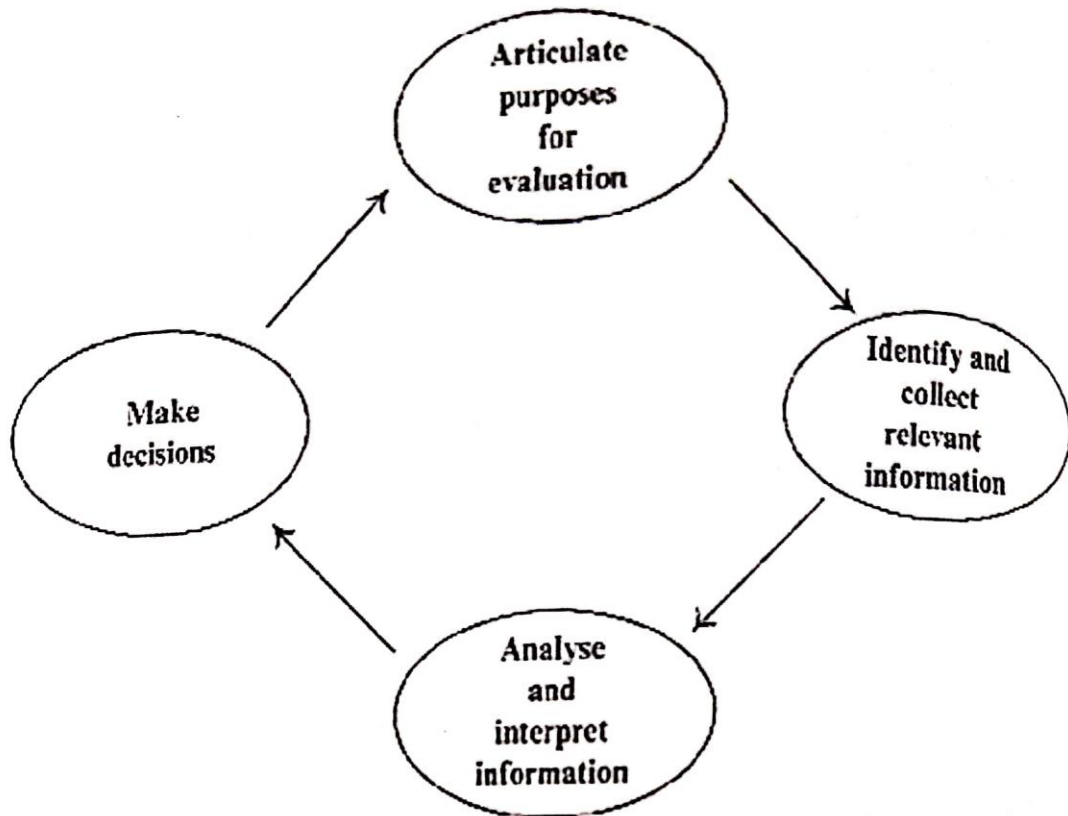
conditions for second/foreign language (L2) learning and, at the same time, to actually notice the value of assessment in language learning.

The goal of this paper is to provide an overview of some key concepts in assessment, discuss the purposes of assessment and the principles and stages teachers need to consider in order to turn assessment into a carefully planned, useful and successful event in their professional practice.

Let us begin by explaining three key terms in the field of assessment that seem to be used interchangeably by teachers (maybe due to the fact that they are translated into Spanish as "evaluación") and which, at times, seemed to mean the same. These three terms are evaluation, assessment and testing. If teachers are aware of the differences that lay beneath each concept, they will be able to better plan and carry out assessment inside their courses.

Evaluation, according to Nunan (1999) can be defined as the collecting of information about aspects of the curriculum which include learners, teachers and materials, and the posterior interpretation of that data with the purpose of making decisions. These decisions could be opening or closing a program, hiring teachers - or firing teachers who do not comply with the standards set by the institution- or certifying a program. Since the evaluation is related to an institution or program, it is usually carried out by external agents who can belong to the government or to private accreditation agencies among others. Evaluation takes time since a wide array of elements needs to be taken into account and it can be carried out in weeks, months or even years.

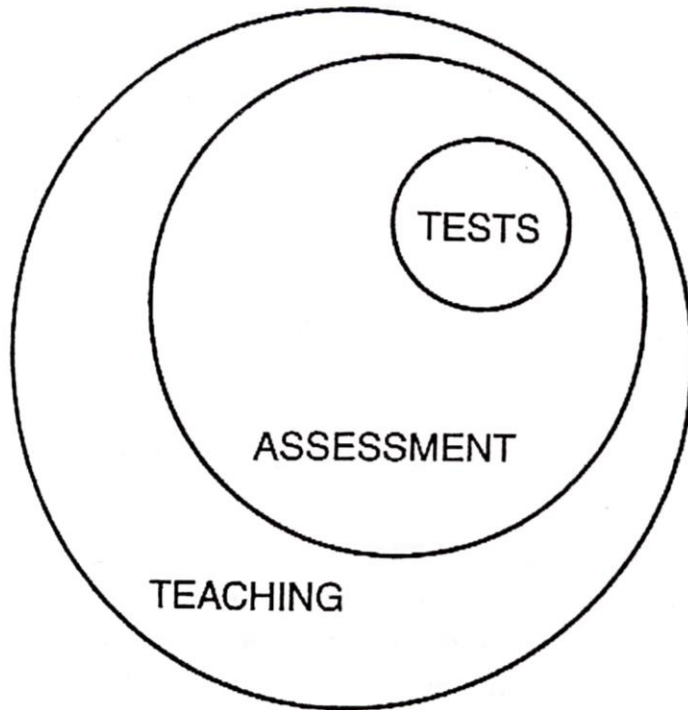
Genesee (2001) suggests that one important purpose of evaluation is accountability (e.g. the process of Acreditación Previa or Registro Calificado in undergraduate programs in Colombia, where programs must demonstrate they have achieved a certain set of standards in order to remain open) and conceives the four basic components of evaluation in a cyclical relationship since they are inter-related and permanent:



Assessment, on the other hand, is for Brown (2004), an ongoing process inside the classroom. He suggests that every time a student answers a question, makes a comment or tries out a new word or a pattern the teacher subconsciously makes an assessment of the student's performance. Brown adds that it is a subcomponent of evaluation since it encompasses what goes on inside the classroom. Here the main agent is the teacher, who designs and applies tasks with the purpose of collecting and interpreting information, this time about what learners can and cannot do as the result of the instructional process that has occurred over a certain period of time. Genesee (2001) explains that "Assessment of an individual student's progress or achievement is an important component of evaluation: it is that part of evaluation that includes the collection and analysis of information about student learning" (p. 145).

Finally testing, as Hedge (2000) suggests, covers the specific steps

that educators or examiners use to measure ability in the language (e.g., a written test on grammar, an oral exam, or a journal to assess writing). For Brown (2004) tests belong to the wider category of assessment which in turn, makes part of the teaching process, as it is evidenced by the following graphic:



He adds that testing refers to the instruments used by teachers in order to assess learners within a particular domain, in this case, language proficiency and defines tests as: "...prepared administrative procedures that occur at identifiable times in a curriculum when learners muster all their faculties to offer peak performance, knowing that their responses are being measured and evaluated" (p. 4).

These definitions provide some insight so that foreign language teachers start reflecting on what they are doing in their classrooms and if those

practices fit the definitions above stated.

It was suggested at the beginning of this paper that in some instances, assessment is something carried out just to "prove" that we were actually teaching and that we have some results to show. It is important then, to establish some academically supported purposes for assessing. Here are some reasons for assessing and although not all of them are ethical, informal talks carried out with different groups of with students point at them consistently.

- To identify how far or how close learners are from the goals established at the beginning of the course, which should be present in the syllabus of the course
- To obtain grades, scores, percentages or other type of information that can be shown to administrators and parents for accountability purposes
- To control discipline problems
- To exert pressure on students exhibiting power and using grades as a tool for obtaining something from them
- To detect progress in the use of the materials selected for the course

Burt and Keenan (1995) put forward these additional reasons:

- To place learners in appropriate institutional levels
- To measure ongoing progress and measure program effectiveness
- To qualify them to enroll in academic or job training programs

Needless to say, FL teachers, and every teacher for that matter, should focus on framing their practice exclusively on academic purposes not only to provide their learners with assessment that displays face validity, but also to avoid assessment being perceived as a process with ethical flaws.

Regardless of the academic purpose teachers have when designing, applying and grading assessment tasks such as tests, portfolios, demonstrations or oral exams, a set of principles needs to be kept in mind while designing these tasks. These principles have to do with time, effort, grading, organization, content, perceptions on the part of students and the re-

relationship between assessment and instruction. Before explaining in detail each one of them, one might ask teachers these questions: When you are creating a test do you consider time restrictions? Length of it? Do you stop to think if your students will perceive the test as useful, organized and reliable? Do you believe that the tests you create might yield similar results when applied to a different group of learners of the same level? Would you include in the test activities which are similar to those carried out in class? Do you have a clear idea of how to score the test in a fair way before starting the first draft? Do you provide your students with clear directions and assessment criteria before hand? These questions will hopefully make you muse about the unconscious process that you as a teacher go through while creating an assessment task

Brown (2004) presents a comprehensive and detailed explanation of the principles that teachers should follow whenever they are facing the task of assessing a group of learners and which address the questions posed before.

The first principle Brown describes is that of practicality. Although it is not defined per se, practicality refers to the cost of the test, the time invested in administering it to your students and also to the time taken to grade it. Tests which are too expensive for teachers or for students are not practical. (e.g., if you create a five-page test and you have to make copies for a group of 45 students). Naturally, the tests being discussed here are those created by teachers, not proficiency tests such as GRE, TOEFL or FCE, whose prices are predictably high. Practicality may also be related to the easiness with which learners can answer it considering the layout or the directions given.

The second principle and maybe the most complex one is that of reliability. Brown (2004) claims that a test is reliable if it is "consistent and dependable". What does this mean for teachers? Consistency means that should the test be applied to the same group of students in different occasions, the test should yield results which are similar to those obtained the first time the test was applied. How dependable a test is will depend on a

number of factors external to the assessment task itself. Brown categorizes these factors as student-related reliability, rater reliability, test administration reliability and test reliability.

Student-related reliability refers to those external elements that can affect a student's performance, thus making the results obtained in the test not reliable. Among these elements we can include tiredness, anxiety caused by the test, being sick or the strategies used by the learner during the test (e.g., some students tend not to read the directions provided in the test and this affects their final score).

Rater-reliability is concerned with the elements that make teachers become less objective while grading an assessment task either by themselves or when they have to grade tests together. Some of the factors that can affect teachers' objectivity are: fatigue, preferring certain students who are perceived as "good", lack of experience in creating and applying tests or a lack of clear scoring criteria.

Test administration reliability relates to the conditions in which a test occurs which can affect the students' performance and ultimately, their score. It is true that sometimes teachers have little power to provide or improve the conditions for taking a test, but in general terms the room should be well lit, noise should be reduced, the test should be legible and always typed, and, if necessary, windows should be open to provide good ventilation. To improve test reliability teachers should not give learners exams which are too long for the allotted time since some items will very likely be neglected by students. Also, tests with ambiguous items or directions or with more than one possible correct answer can and possibly will affect the reliability of the score obtained.

The third principle that teachers and examiners should consider when creating an assessment task is that of validity. Brown quotes Gronlund (1998) who defines it as "the extent to which inferences made from assessment results are appropriate, meaningful, and useful in terms of the purpose of the assessment" (p. 22). In other words, if a test is valid it should measure what it intended to measure in the first place. For instance, if the

purpose of a test is to assess vocabulary and the teacher creates an assessment task focused on grammar, the validity of this instrument is obviously lost. As with reliability there are different categories that can be related to the fulfillment of this principle. Brindley (2001) suggests that first of all, a test is valid if it models the topics or matters about which teachers expect to draw conclusions, and if it entails performing the measured behavior on the part of learners. Another piece of evidence that reveals the validity of a test is to determine if the criterion has been attained, which means that learners could be assessed externally and they would consistently exhibit a certain behavior (e.g., pronouncing correctly some sounds, conjugating some verbs accurately). A final consideration is whether the test is perceived by the examinees as relevant, just, and useful for enhancing learning, that is, if the test has face validity. Face validity is present in an assessment task if the format is organized and displays tasks previously carried out, if it is feasible to complete it within reasonable time limits and if tasks are related to the course materials and topics.

Authenticity is next on the list of principles and it may be difficult to define and to achieve given the fact that teachers in Pasto are immersed in an EFL setting where classes tend to be short, decontextualized from real-life needs and conducted mostly in Spanish. If teachers want to bring authenticity in their assessment, tasks have to be closely related to what learners might be expected to do in a real-life situation. This is easier said than done since teachers might create a task considering what they think learners will have to say or do in a context where the L2 is spoken or using vocabulary exclusively from the textbook. In both cases authenticity is lost. A good idea is to rely on a colleague who can proof-read your draft and give you some insights on whether or not the language of the test is natural and if items are contextualized and feasible.

The final principle and, dare we say, the most neglected on the part of teachers is that of washback. Washback is simply the effect a test has on instruction and consequently on learning. An example of washback can be the ECAES in Colombia. After the first ECAES were administered some

programs started special courses whose purpose was not to enhance learning, but to teach for the test so that the next generation of test takers achieve higher scores. This illustration shows the effect of a test (ECAES) on teaching and learning (the creation of additional courses). Certainly washback is not negative and it can provide teachers with useful information on what are the learners' strengths and weaknesses; since assessment is an ongoing process, instruction should be modified halfway to address difficulties and get students back on track. Washback can also inform teachers about the effectiveness of certain assessment tasks and directions, and it can also serve as a guide for learners, some sort of self-assessment which will help them identify their own strong and weak areas in a course. Washback can be provided right after handing the tests back to students. Afterward, the teacher can comment on the test, give learners the right answers or help them recognize what they did right or wrong.

All these conceptualizations and definitions of terms are not useful unless teachers put them to good use and the only way of doing this is by creating assessment tasks that entail all these ideas. Theory is only valuable when it informs the professional practice of teachers and practice is made more efficient through the knowledge of theory. However, designing assessment tasks, and more specifically tests, may be a daunting experience for novice teachers who discover on a Sunday night that the test they created for Monday morning has 36 items of the same type and making the items fit to a final score of 5.0 is even more difficult than coming up with all those questions. In order to avoid this type of practicality nightmares it is important to keep in mind that assessment needs planning, as virtually every activity of teaching does. Brown (2001, 2004) suggests a checklist of steps that can be used to make sure assessment is biased for best.

The first thing is to define why you are applying the test or task and what is it that you want to test. Your objectives will be easily set up if you have an organized syllabus with the main aspects of the course, that is, with what your students should be able to do by the time assessment takes place. After this has been decided, you turn to creating test specifications; (i.e.,

time for completing the test, number of sections the test will have, topics and types of items to be included, skills that will be assessed). The third step is to write a draft of the test which will allow you to judge if it is valid and authentic and if the scores assigned to each item are right. This draft must be revised to check if directions are unambiguous, if the language is clear and to avoid typos and grammar mistakes. Once the draft is revised, and if possible, someone else might take a look at your test and give it some feedback. Finally, the test must be applied considering that it should be followed by feedback on your part and on the part of your students. This washback should then be reflected on instruction and the cycle starts again...

As it can be seen, evaluation, assessment and testing, play an important role in foreign language teaching and learning, and it is time that teachers pay more attention to the assessment process in order to bring about important transformations. Nowadays students have to face standardized assessment, standardized exams (TOEFL, ICFES, PET, FCE) and performance-based assessment. In order to be able to face these challenges, learners need to be given the opportunity of experiencing adequate and interconnected learning and assessment practices inside the FL classroom. Assessment needs to be freed from that secrecy aura some teachers want to confer to it since pretty soon external agents will be in charge of assessing our students and, ultimately, it will not matter if you are a good, bad, nice, demanding or a lazy teacher; all that will matter is that your students are able to perform efficiently in the L2 in diverse situations. Are we up for that challenge?

REFERENCES

- Brindley, G. (2001). Assessment. In R. Carter & D. Nunan (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp.137-143). Cambridge: Cambridge University Press.

Brown, H. D. (2001). *Teaching by principles*. White Plains: Prentice Hall Regents.

Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. White Plains: Pearson Education.

Burt, M., & Keenan, F. (1995). *Adult ESL Learner Assessment: Purposes and Tools*.

The National Clearinghouse for ESL Literacy Education. (ERIC Document

Reproduction Service No. ED 386962)

Genesee, F. (2001). Evaluation. In R. Carter & D. Nunan (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp.144-150). Cambridge: Cambridge University Press.

Hedge, T. (2000). *Teaching and learning in the language classroom*. Oxford: Oxford

University Press.

Nunan, D. (1999). *Second language teaching and learning*. Boston, Heinle & Heinle Publishers.