

1. Título de la ponencia: **IN-Service Teachers' Metamorphological Process in Evaluating English as a Foreign Language in Costa Rica**
2. Ponentes: M.A. Gustavo Álvarez Martínez
M.A. Mariela Porras Núñez
B.A. Deyanira Cubillo Ramírez
M.A. Nandayure Valenzuela Arce
3. Dirección: Heredia, Costa Rica.
4. Institución: Universidad Nacional de Costa Rica (UNA)
5. Grado académico: Profesores de Inglés de la Escuela de Literatura y Ciencias del Lenguaje de la Facultad de Filosofía y Letras.
6. Correo electrónico: lgtavo@yahoo.com; emaryp@hotmail.com;
deyacubillo@gmail.com; nandayurevalenzuela@ice.co.cr

In-Service Teachers' Metamorphological Process in Evaluating English as a Foreign Language in Costa Rica

In Costa Rica, the Ministry of Public Education has grounded English learning programs on the Communicative Language Teaching Approach (CLT). It was thought that at the completion of high-school, the adoption of this methodology would enable students to communicate effectively in authentic contexts. Although a great number of in-service foreign language (FL) instructors master CLT procedures, most of them have very little or no formation on a compatible evaluation. Consequently, countless language teachers continue using traditional evaluation to test the CLT authentic outcomes.

It is a fact that traditional exams generally test grammatical forms or vocabulary items. These non-authentic kinds of tests just measure the students' ability to choose correct alternatives to get a good grade rather than using those grammatical structures in natural oral contexts. In this regard, Hughes (1995) points out that, "...if the test content and testing techniques are at variance with the objectives of the course, then there is likely to be harmful backwash"(p. 1).¹ Then, designing and administrating non-accurate tests can have a negative impact on the learners. Students' learning is focused on short term memory because traditional tests emphasize the evaluation of knowledge on language structures and vocabulary; in these tests, very little (if none at all) attention is given to the use of English in real conversations and-or written transactions.

Similarly, O'Malley and Pierce (1995) debate that there are two major issues that language instructors should pay attention to: one is that "current assessment procedures do not assess the full range of essential students outcomes and that teachers have difficulty using the information gained for instructional planning" (p. 2). Since plenty educational programs of the country share this broad evaluation

constrain, it is imperative to devise a solution. As long as FL instructors continue administering traditional tests, the number of English proficient students will continue to be very low. The instructors must consider whether non-authentic tests are accurate reflections of real language use, or if they want the high school population to achieve higher language proficiency when they finish their studies.

Finally, if there is a negative backwash effect on Costa Rica's teaching and learning process, and a mismatch between the "National English Plan" objectives and the kind of exams applied to value the students' language knowledge, thus, there is an urgent solution to approach.

The essentials of language teaching are to convey answers to any of the problem that occurs on the students' learning process. For that reason the program of the Master's in Second Languages and Cultures (MSLC), of the Escuela de Literatura y Ciencias de Lenguaje (ELCL), at the Universidad Nacional (UNA) launched a pilot advanced training for specializing instructors on alternative evaluation, and the team involved (trainer and trainees) carried out a dual action research. The two research perspectives contributed to derive evaluation theory that can be pertinent to the country, and to prepare more qualified instructors in authentic evaluation, to respond to prompt solutions of the problem stated.

For the sake of clarity, from here on we refer to the different academic populations as stated in the table below.

Table 1. Description of Academic Groups at the ELCL

Category	Acronym	Description
Trainers	(TRs)	Instructors leading the advanced training for specialization.
Trainees	(TEEs)	Participants of the advanced training for specialization.
Master degree instructor	(MDI)	Masters program educators.
Graduate Students	(GSs)	Masters program learners.

Holding Back the Fruition of Alternative Evaluation

Worldwide Communicative Language Testing is one of the areas in Foreign Language Teaching (FLT) that is still in its maturation development, with scanty professionals prepared in this field. Hughes (1995) confirms that “too often language tests have a harmful effect on teaching and learning; and too often they fail to measure accurately whatever it is they are intended to measure”(p. 1); in other words, teaching methodologies are inefficient without the implementation of theoretical and practical well-grounded evaluation.

There are three main causes that can be accounted for the country’s FL learners’ weak development of English for real communicative purposes. One can be attributed to the lack of competence that many in-service English teachers have in alternative evaluation. Another is the deficient guidance that examining boards in public and private schools give to language teachers. And the lack of opportunities that English professionals have to develop expertise in alternative evaluation due to the deficient offer of this type of specialized learning in national academic programs or advanced trainings projects.

Evaluation Shortcomings, Some Venues to Improvement

Most teachers who lack knowledge in evaluation tend to do their best in designing and administering their test but fail in doing it properly. This is simple because each testing situation is unique and the tester must be equipped with the appropriate evaluation knowledge, skills and techniques. This evaluation weakness in designing and administering exams evidences the learners’ failure in language acquisition and thus their low language proficiency in communicating. As a solution, Hughes advises that language instructors should develop tests which should be valid and reliable.²

Also, teachers who do not have testing skills cannot dispute the school authorities about the irregularities that they perceive in evaluation. Inexperienced instructors must follow already fixed patterns from the school evaluation system if they want to continue working in the educational institutions. If the GSs show well-designed tests and the learners' progress scales to support the urge for making changes, principals and coordinators of language institutions and schools can study the proposal and consider modifying institutional evaluation parameters. Positively persuading other professionals and school leaders is the key to succeed in applying authentic evaluation in any group, class, time and place.

In both public and private institutions, there is an all-purpose examining board that regulates all the standard subject matter tests of the different school programs of the country; very often, however, these examining boards are not integrated by teachers endowed with expertise in L2 alternative evaluation. In most of those schools, the evaluation teams equally approve or disapprove the foreign language tests as they do with those of academic-subject matter. Analyzing the English tests submitted for approval using the same parameters as of those applied for the rest definitely disrupts the students' learning process for acquiring the four English skills, because teachers are forced to follow the other academic tests' patterns. Language instructors lack evaluation knowledge to strongly argue against harmful testing practices. This evident deficiency in the examining boards' procedures and the teachers' lack of knowledge in testing make the English language evaluation system in all institutions to work incompetently. Examining boards should incorporate a high-skillful language instructor who can easily advise FL inexperienced evaluators and support those who are more advanced. Only by this means, alternative evaluation can be adopted in both public and private schools' programs.

To satisfy the nation's need of skillful instructors in Authentic Evaluation, the program of the Master's in Second Languages and Cultures (MSLC), of the Escuela de Literatura y Ciencias de Lenguaje (ELCL), Universidad Nacional, includes four courses³ that focus on authentic evaluation. Although all of them are valuable, the course *Evaluation and Measurement* centers on having graduate students acquire the knowledge and skills needed to carry out properly evaluation practices (design of test and rubrics, test administration, evaluation research, and others) and to assess the FL learners' language skills in a more efficient way.

Internal Approach of Realities

The formation of professionals in the alternative evaluation discipline is a key contribution of the MSLC to the country's nationwide bilingualism challenge; yet, the program cannot train, in the short time required, the myriad of English teachers of the country who lack instruction in authentic evaluation. Although the GSs of the Evaluation and Measurement course at the MSLC extend their knowledge-base and develop skills in authentic evaluation, they do not attain a specialization level as to be sufficiently competent to provide training in this area to other instructors.

In view of the limitation stated above, the MSLC conceived an alternative option; the masters program took action implementing a pilot specialization project to prepare a triad of professors in authentic evaluation. The pilot project's effectiveness was tested carrying out research in the MSLC's *Evaluation and Measurement* course.

The TEEs carried out a dual action research to endure a new academic experience. Through the internal approach of realities, the GSs nurture the TEEs with their insights and reflection about evaluation once they acquire knowledge in this specialized discipline.

From the Caterpillar to the Butterfly

Through an introspective and extrospective methodology, the GSs' of the Evaluation and Measurement course examined their own experience as foreign language teachers before and after acquiring the theory taught in classes. The 2009 promotion evidences through academic products that they have certainly evolved as true testers and gained in depth the necessary evaluation theory to spread their knowledge to others. This metamorphological process could be feasible for the book used in the course *Testing for Language Teachers* by Arthur Hughes and the MEI and TR's guidance. Undoubtedly, Hughes's written assistance was an excellent tool that provided direction to the GSs. Next to the expert TR's guidance, the GSs designed and administered the four English skill tests, listening, reading, writing and speaking. Conjointly, the GSs analyzed and analyze their products to nurture the other GSs and came out with possible solutions or changes to the problems they encountered. This pedagogical context had a critical and insightful atmosphere not only for the GSs to grow but for TEEs to treasure academic experience.

Master's Candidates' Voices

A transformational process was evident in the 2009 GSs. They enlightened each other and the TEEs of the specialization project with their insights that were recorded in academic products and in the researchers' notes.

The enormous contribution that the GSs gave to the TEEs really allowed the specialization project members to discuss and analyze the students' theorization and support the importance of action research for certified but inexperienced instructors. Trainees can rapidly gain experience in a classroom where a senior professor is guiding a group of teachers.

For the sake of clarity, from here on we refer to the MGSs who served as informants to this action research as shown in the table below:

Table 2. Coding informants' names, GSs from the Evaluation and Measurement Course

Category	Acronym For Informant	Description
Graduate students (GSs)	GS A	Informant from the Evaluation and Measurement course
	GS B	Informant from the Evaluation and Measurement course
	GS C	Informant from the Evaluation and Measurement course

Through the voice of students in the Evaluation and Measurement course, the TEEs were enlightened of some evaluation drawbacks. A sample of three clue informants was selected to portray the GSs' experiences, views and reflections of the course. This internal, personal perspective is portrayed as follows.

GS A Informs

GS A speculated about how she and her classmates, as in-service English teachers, could discern if they were acquiring enough knowledge on alternative evaluation as to propitiate with their testing practices a positive backwash effect on their students. She wrote that the only way to find out was by replicating over and over in their class, the efficient testing practices exemplified in the masters' class, which were grounded on alternative evaluation theory. GS A stated that it was possible to embark "on the task of applying a different approach of teaching, testing and assessment following the guidelines of the *Common European Framework*, on one hand, and Hughes's text *Testing for Language Teachers* on the other. In addition, GSs could rely on the guidance of their Testing and Measurement course professor."

GS A explains that at the beginning of her transformation process, from a teacher with very basic knowledge of L2 evaluation to the development of competence as an alternative evaluation tester, made her feel sad because she contributed, although not aware of it, to limit her students opportunities to learn English for real world purposes. GS A explains that in her high school, where she taught 11th grade students, she was forced to teach them only strategies and models included in the textbooks, which were aimed at preparing learners to take the TOEFL⁴ and TOEIC⁵ exams; this teaching approach raised students' anxiety and stress. Both of the appointed tests are classified as High Stake assessment, these are tests (often standardized) associated with high stakes decisions such as employment, graduation or access to further education. However, GS A acknowledged that she would have preferred to create a communicative environment for the students to acquire the L2 and motivate them to interact successfully in English.

In many EFL contexts, English is a mandatory subject in the curriculum. In order to be admitted in colleges and graduate schools, students must demonstrate English proficiency at particular levels which differ from setting to setting."⁶ GS A had to pay a lot of attention in building up a fifth and sixth language skill (intuition and fear) on students to pass these two high stake assessments. GS A wondered how eleventh graders could ever be able to authentically communicate in English if they were just able to mechanically answer already fixed responses.

GS B Informs

GS B affirmed that her knowledge base and skills were on authentic evaluation were enhanced mainly by means of implementing theory on the subject. She designed and administered tests to evaluate the four language skills learners of

a private language institution where she works. In fact, this GS made changes in the way evaluation was administered in this language institution. She pointed out that the design and administration of the listening and writing tests were “the most fulfilling experience of all the tests I have implemented because these language skills are being assessed incorrectly at FPRS,⁷ and I feel responsible for that.” GS B realized that the listening exams were non-authentic task; all items were multiple choice.

After enhancing her knowledge on alternative evaluation, Gs B immediately included the necessary *fresh starts*⁸ to have a positive backwash effect on students as Hughes advices. Additionally, the writing test models lack an important element in testing. Instructors do not use scoring scales to subjectively evaluate the writing products of the learners. GS B eagerly made a change, she designed the corresponding rubrics to the groups she was teaching and an authority asked for revision and inclusion of the element in their evaluation system. Gs B turned into a significant social agent who started repairing and improving the institute English teaching.

GS C Informs

GS C taught at a private elementary school and she also belonged to the examining board of the same educational institution. Before her evolution in authentic evaluation, she did not know that when the objective was to assess the students’ communicative competence, language tests were designed and administered differently. Hughes (1995) points out to the difference emphasizing that “language abilities are not easy to measure; we cannot expect a level of accuracy comparable to those of measurements in the physical sciences.” GS C herself destroyed others teachers’ language tests because the members of the group used

the same paradigms of regular academic subject matter such as mathematics, science or social studies. The evaluation team incorrectly used and emphasized elements on language tests that washed back negatively the classroom instruction. The objectives and the tasks of the tests did not match with the teaching strategies used to develop the oral skills, for example. Having an ample knowledge in testing, the graduate student could contribute and make a change in the board inefficient procedure. She insisted that from now on, language teachers will have a one hundred support to achieve their teaching expectations.

Overall, the voice of the 2009 graduates, full of awareness on evaluation shortcomings, will be heard and followed. Their experiences can serve as an example of the need to move from the blind stage of ignorance in which they once were, towards the stage of professional growth attained by the end of the *Evaluation and Measurement* course.

From the Perspective of the TEE

There is a noticeable evolution on the TEE as a result of the experiences provided by the pilot training project on evaluation. The specialization candidate was once a graduate student in the same evaluation course and had to undergo the same academic journey. Next to this educational learning in testing, the researcher could observe the transformation of the group under study. The GSs metamorphological process lead TEEs to study even more in depth different theories and approaches on evaluation. GSs' insights showed that many public and private school evaluation inconsistencies can be solved. The GSs unveiled many of those problems and implemented efficient strategies to defang their effects. However, there are still many English teachers in the country who lack knowledge in evaluation, so the negative

backwash effects of deficient or incongruent evaluation keeps holding back, nationwide, many students' goal to become bilingual.

Further investigation should be carried out in order to get a wider panorama of the phenomenon and to device more alternatives for in-service English teachers to learn the ways of alternative evaluation.

REFERENCES

Hughes, Arthur. (1995). *Testing for Language Teachers*. 7th ed. Cambridge:

Cambridge UP, 1995. Print.

MEP, COMEX, MEIC, INA, CINDE, y Estrategia Siglo XXI. (2007). Plan Nacional de

Inglés. San José, Costa Rica.

O'Malley, Michael and Lorraine Valdez. (1996). *Authentic Assessment for English*

Language Learners: Practical Approaches for Teachers. Reading, MA.:

Addison-Wesley.

FOOTNOTES

¹ A negative effect of testing on teaching and learning (Hughes, *Testing for Language Teachers*)

² A test is said to be valid if it measures accurately what is intended to measure and a test is reliable if the scores obtained by a particular student are similar.

³ Testing and measurement LPE 705; Assessment LPE 707; Evaluation of curriculum LPE 7310; and Design and evaluation of didactic materials LPE 714.

⁴ The *Test of English as a Foreign Language* (TOEFL) measures the ability of nonnative speakers of English to use and understand North American English as it is spoken, written and heard in college and university settings.

⁵ The *Test of English for International Communication* (TOEIC) measures the everyday English skills of people working in an international environment.

⁶ A Close Examination of High Stakes Assessments at
https://oncourse.iu.edu/access/content/user/mikuleck/Filemanager_Public_Files/EFL_Assessment/Unit_2/Unit_2_Overview.doc

⁷ Acronym of the educational institution in which this graduate student works. The full wording is not provided to protect the institution image.

⁸ Different parts on a test to avoid only one task.