


1. Título de la ponencia: **Testing and Measurement as a Multipurpose Mirror: the Need to Ignite Introspection in Language Teachers** 
2. Ponentes: M.A. Mariela Porras Núñez  
B.A. Deyanira Cubillo Ramírez  
M.A. Gustavo Álvarez Martínez  
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: [emaryp@hotmail.com](mailto:emaryp@hotmail.com); [deyacubillo@gmail.com](mailto:deyacubillo@gmail.com);  
[lgtavo@yahoo.com](mailto:lgtavo@yahoo.com); [nandayurevalenzuela@ice.co.cr](mailto:nandayurevalenzuela@ice.co.cr)

### **Testing and Measurement as a Multipurpose Mirror: the Need to Ignite Introspection in Language Teachers<sup>1</sup>**

In Costa Rica, some efforts have been made to reorient the path of Foreign Language (FL) teaching to yield a more successful outcome; for instance, the national program Costa Rica Multilingüe,<sup>1</sup> which adopted the Common European Framework of Reference (CEF) as the taxonomy to measure language ability in Costa Rican public and private institutions. Unfortunately, the evaluation of the communicative objectives proposed in the policy of the Ministry of Education continues to be based on a quantitative perspective that tries to determine how much students know rather than how well they can actually use the language to express ideas and convey meaning, as the foremost objective of communication.

In view of that panorama, it is imperative to ensure the implementation of non-traditional evaluation based on real communicative principles. Nevertheless, most language instructors are not prepared to apply evaluation as a more qualitative approach. In this regard, O'Malley and Valdez (1996) point out that notwithstanding the high amount of professional time spent in assessment-related activities, "pre- and in-service programs to date have not familiarized teachers with issues in authentic assessment, nor have they prepared them to design and use this type of assessment for instructional planning" (p. 6). More than a decade later, the truth underlying such affirmation is still evident. Therefore, training on authentic evaluation becomes a must.

Language instructors need to undergo the route of introspection, where they can reflect and make pertinent decisions about their role as teachers and evaluators. One of the failures of the educational system in Costa Rica might have been to expect EFL

---

teachers to know how to become successful language evaluators on their own. Nonetheless, for educators to succeed in this endeavor, they require real guidance through formal instruction.

Precisely, one of the few grade programs that serve as a vehicle for professionals to improve their daily evaluation practices is the Masters in Second Languages and Cultures at the Universidad Nacional. A highly salient component of the masters' program is the *Evaluation and Measurement* course, which provides masters program learners (GSs) with the first encounters with the area of alternative evaluation. Considering the pertinence and enriching nature of this particular course, it was selected as the basis for the first Pilot Specialization Project on Evaluation,<sup>2</sup> undertaken by three professors from the ELCL<sup>3</sup> during the second semester of 2009. Guided by the trainer instructor leading the advanced training for specialization (TR)—also coordinator of the masters' program and professor of *Evaluation and Measurement* course—the trainees (TEEs) carried out action research taking advantage of the interaction that takes place in a classroom where masters program learners are being nurtured with the principles of authentic evaluation.

For the sake of clarity, in this document 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
<b>Trainers</b>	<b>(TRs)</b>	Instructors leading the advanced training for specialization.
<b>Trainees</b>	<b>(TEEs)</b>	Participants of the advanced training for specialization.
<b>Master degree instructor</b>	<b>(MDI)</b>	Masters program educators.
<b>Graduate Students</b>	<b>(GSs)</b>	Masters program learners.

Two main aspects motivated the participants of the advanced training (TEEs) to follow the path that qualified senior professors have cultivated and advanced for years. One is the national reality in the field of evaluation in second languages in Costa Rica, and the second is their personal interest in specializing and acquiring expertise in that fundamental area of applied linguistics.

Through commitment and cooperative work, it became plausible to analyze the dynamics of the aforementioned course in order to obtain valuable insights pertaining to the area of evaluation. With the aim of obtaining more objective outcomes, the research was directed from two perspectives. Part of the analysis was carried out from the GSs's perspective,<sup>4</sup> considering their opinions, their class performance and other relevant information depicted in the different assignments handed in. The other part of the scrutiny did not take into account the learners' point of view. Instead, it was completed from an external standpoint.

In the following pages, the reader will be exposed to the account of the second examination. First, there is a brief description and rationale of the precise way in which the outside perspective was addressed. Afterward, there is an analysis of classroom interaction under the light of theoretical principles along with the insights provided by the professor responsible for teaching the masters' course. The final part of the document discusses the impact of this action-research on the pilot specialization project in evaluation. In addition, this same section reveals how the aforementioned research approach manifests itself as a crucial initiative to train the vast of knowledgeable faculty of the ELCL in different branches of applied linguistics, thereby benefiting all: the more experienced instructors, the less experienced instructors, as well as the novice instructors who are initiating their professional endeavor.

### **External Inspection of the Truth**

The action research described here is characterized by the role of the participants of the advanced training for specialization (TEE) as an outsider, i.e. as a non-participant observer whose analysis is based on the interaction between her own insights and the ones provided by the professor of the course, supported by pertinent theory. The systematization of data gathered from the observations followed a cyclical procedure including analysis of theory, critical discussions, as well as proposal and implementation of strategies to address weaknesses and to enhance strengths depicted.

### **Observation: A Critical Spyglass**

Notwithstanding the type of research being conducted, observation stands as an essential data collection instrument, which must be justified by a clear purpose and which requires responsibility and objectivity.

In this particular case, observation was a necessary data gathering instrument aimed at projecting classroom reality as it occurred and thus, at providing clear evidence to determine how the *Evaluation and Measurement* course contributes with the formation of pioneers in authentic evaluation, and the succeeding consequences of such accomplishment for educational policies at the national level.

### **The Multipurpose Mirror**

By means of non-participant observations, it became evident how the GSs' evaluation system was positively influenced along the course. The methodology based on the notion of guidance, together with the theoretical approach followed, and the type

of assessment applied made the class a multipurpose mirror that awakened and/or reconfigured the concept of introspection.

### **The Fine Line between Guidance and Prescription**

Learners at any age are no longer seen as passive receivers of knowledge, and the idea of instructors owning the truth has been rejected during the last decades. Resembling Paulo Freire's critical pedagogy,<sup>5</sup> the *Evaluation and Measurement* course provided GSs with a very active role throughout the whole process. They were empowered to establish purposeful connections among theory, reflection and action taking. It is fundamental to draw attention to the fact that the fine line between prescription and guidance was well delimited.

While there were countless sources of oral and written guidance for each task assigned in and out of the class, there was no evidence of prescription whatsoever. On one hand, GSs were given extensive guidelines explained in detail as to set the standards of performance and the requirements in terms of format and content structure. On the other hand, guidelines asked and encouraged those learners to make of each product a realistic representation of themselves and their in-depth interaction with theory. As an illustration, part of the instructions for the testing portfolio—one of the most important requirements of the course—stated,

Along the course, you were testing Hughes' and your own hypothesis and theories with regards to the evaluation of language abilities, so now you should be ready to present a much more solid position (in favor or against) the design, administration, and backwash effects of communicative

evaluation and rating scales to measure productive language skills (Final guidelines for the testing portfolio).

It is clear that GSs were not told what ideas to write or which theory to agree with. Instead, they were given the opportunity to gain practical understanding through the critical interaction of their prior knowledge and theory studied. The main purpose of the type of guidance implemented was to obtain long-lasting results. According to Kirschner et al. (2006), that can aid learning “to be indeed a change in long-term memory” (p. 75).

From lesson one, the masters program instructor (MDI) responsible for the course served as an example of organization and commitment, and her role as facilitator was easily depicted. The class agenda, for instance, clearly stated a careful planning including the projected active participation of GSs in different learning experiences. The following task extracted from one of the agendas may serve as an illustration, “Deriving the learning gained from Hughes’ text on testing writing *by analyzing two tests.*”<sup>6</sup> Although GSs were guided throughout the discussion, they were not prescribed as to provide specific answers or to agree with pre-established ideas.

The course objectives were addressed in accordance with the academic standards of the masters program, the MDI’s teaching philosophy and the support of relevant theory that allowed GSs to acquire and put into practice major principles of alternative evaluation.

According to the masters program, the *Evaluation and Measurement* course approaches communicative evaluation of the target language and provides learners with the opportunity to study and interact critically with the theory as well as to implement—in their educational praxis—the propositions addressed. It is essential to point out that measurement involves scrutiny of final products and tests designed under the principles

of alternative evaluation. This assumption purports that GSs learn how to depart from instruction and their pupils' needs when constructing communicative items, which must also resemble real and meaningful contexts.

What was observed in class throughout the course was congruent with the MDI's teaching philosophy. As claimed by this instructor, for the learning process to be effective, learners must be guided and challenged at the same time. Her philosophy is based on the *Challenge Approach*,<sup>7</sup> which states that with appropriate guidance, pupils are led to the path of responsible risk-taking. In this regard, Alvarez et al. (2008) claim that learners "must take responsibility for using provided learning tools efficiently; select strategies that help them work better; and risk to set ambitious-but attainable-personal learning goals through self-challenge" (p. 27).

It is clear how the dichotomy between *telling learners what to do* and *having them make decisions on their own* was addressed through challenge and sufficient guidance.

When asked about the achievement of objectives, the professor of the course under investigation stated that, "In spite of teacher and students' efforts to learn and apply theory on evaluation with 100% success, the master-learners' efficient acquisition and performances varied in degree of proficiency" (Valenzuela).<sup>8</sup> In addition, she recognized that such variance was due to the expected dissimilarities among learners, in terms of the cognitive and emotional mechanisms to which they resort along the process.

The diversity described was also perceived by the TEE, who witnessed how learners dealt with theory during discussions, mastered metalanguage and were able to extrapolate the acquired knowledge to the various assigned tasks, in different moments and at their own pace. The philosophy adduced by MDI was also evident in this sense,



given that GSs were challenged to discover their own weaknesses and to look for strategies to overcome them. By participating in class tasks such as *solving testing problems* and *deriving theory by analyzing tests*,<sup>9</sup> implementation of theory was put into practice; and at the same time, cooperative work was enhanced by students' sharing ideas and giving reciprocal feedback.

As a case in point, during the class assignment called *analyzing theory on testing techniques, stages of test construction, and testing listening*,<sup>10</sup> a half of the class was in charge of writing the guidelines to design a listening test while the other half was responsible for stating the guidelines to administer the test appropriately. Certainly, the outcomes may have not demonstrated expertise in the topic studied; however, they showed GSs' awareness of their own and others' limitations in addition to openness to listen and learn from others.

### **Leading Leaders**

As supported by evidence, GSs were guided throughout all the process and were challenged to carry out in depth analysis of the theory and introspection of their role as evaluators. That is a significant aspect when considering the kind of critical and reflective practitioners needed by Costa Rican educational system. Hancock (1994) concurs with Allwright (1988) who argues that "greater quality of learning can be ensured by putting the control over learning in the place where the learning is occurring, namely in the mind of the learner" (p.3).

In the course inspected, there was also space for action research intended to lead the GSs to the achievement of the course objectives and the improvement of their own teaching practice. In this way, these future MDIs were gradually and implicitly

invited to become leaders, who would make a difference in the approach of authentic evaluation and measurement within their work context.

A major remark of the *Evaluation and Measurement* course is the design and implementation of grading scales (rubrics), for it represents one of the main components of the authentic non-traditional evaluation practice pursued nowadays. As supported by Mertler (2001), "Scoring rubrics provide at least two benefits in the evaluation process. First, they support the examination of the extent to which the specified criteria has been reached. Second, they provide feedback to students concerning how to improve their performances (n.p.). As a matter of fact, the evaluation applied during the course was congruent with the evaluation principles studied. The MDI evaluated GSs' performance with rubrics that she carefully designed in accordance with the objectives and characteristics of the course. Specific scales were used for each product (introspective essay, lexical portfolio, etc.) and, although only once, a self-assessment rubric was applied for GSs to assess their own work.

This was highly beneficial because these learners were able to draw both on the theory studied and on the professor's example when creating rubrics. Additionally, they were provided with samples of scales designed by students from previous years, which gave them a wider and clearer perspective on the demanding task.

At first, the TEE perceived some reluctance or fear from GSs with regard to the use of grading scales, justified perhaps by the innovative nature of the practice and the possible resistance that it could produce in their colleagues and even in their school authorities. However, the analysis of theory and the practical experience guided by the MDI, gave GSs the confidence and the foundation to believe in and to start the process

of learning to design valid rubrics to evaluate their own pupils. As expressed by the MDI of the course,

There was great improvement of the master-students' evaluation performances in real life situations and this created a positive backwash effect in their teaching praxes. The pupils realized that the testees' final outcomes had been properly or improperly oriented during the learning process and they (master-learners) prompted to modify those of their teaching performances that had had a bad effect on final outcomes (Valenzuela).

Undoubtedly, GSs became pioneers<sup>11</sup> who were given a point of departure to make necessary changes in their professional context, and to ignite in their colleagues the desire and need to scrutinize and improve the way in which evaluation of EFL communicative competence is approached.

### **Targeting Processes of the Heart and of the Mind**

From an outside perspective, the TEEs are confident to say that learners were seen as integral beings, and as social agents who have a voice, not only as students from the masters' program but also—and mainly—as professionals who are willing to eradicate unfruitful evaluation practices in order to implement appropriate ones. Such demanding task is consistent with another core objective of the CEF, i.e., to foster reflection on current teaching and evaluative practices in order to ensure that the real needs of the learners are adequately and responsibly met (p. 1).

GSs' voice was always recognized as a way to trigger criticism and introspection. The dynamics of interaction were characterized by constant feedback from the MDI to

the students and vice versa, as well as among learners. GSs always had the opportunity to express ideas, ask questions, clarify doubts, and share experiences. They were not only expected to gain understanding in the area of evaluation, but also to extrapolate their knowledge and act as social agents able to promote necessary changes in their educational context. In this sense, the MDI asserted,

The voice of the master-students and specialization candidates was, and is, key feedback for me. I was always open to hear the course participants' critics and use their argumentation to improve my knowledge, skills, materials, attitudes, etc. The adjustments were used in behalf of the students' learning and to better my own performances in class (Valenzuela).

It is discernible that the congruency between the academic goals stated in the program under pertinent theory to the field of authentic evaluation and the MDI's teaching philosophy turns the course into a multipurpose mirror through which GSs are able to see beyond their own image as language instructors.

### **Advocating for Specialization**

A relevant component of the advanced training for specialization was the participation of TEEs in the class dynamics by teaching in one of the lessons, as proposed by the TR. From non-participant to participant observers, TEEs had a short but valuable opportunity to emulate relevant elements observed as a way to put them into practice and analyze similar or new outcomes.

Via theoretically grounded tasks, TEEs encouraged GSs to apply the principles of evaluation on the basis of critical reflection.<sup>12</sup> For instance, when analyzing cartoons that

debated flaws in evaluation praxes, GSs were able to establish connections between the intended message in the caricatures and the reality of their own working context. Furthermore, GSs showed to be more acquainted with the fundamentals of authentic evaluation, which led to more explicit concern and willingness to accept the challenges that their new knowledge brought on them.

That positive outcome was also indentified by the MDI, who—by the end of the course—expressed,

The master-students also realized that they are pioneers on the authentic evaluation practice, and that it is their responsibility to spread this fair measurement trend. They admitted being somewhat scared of the challenge for social change that their new knowledge posed on them, but they also seemed committed to carry out the task” (Valenzuela).

Another crucial phase of the action research was the in-depth discussion and analysis of input gained throughout the study—including observations, discussions, class teaching, etc. Frequent meetings were scheduled with the trainer instructor (TR) with the purpose of reporting outcomes from the observations. The insights were always supported by and complemented with the examination of theory of prestigious scholars others than the author of the textbook used by GSs in the course.

Both TEEs and TR were able to depict weaknesses and strengths of the *Evaluation and Measurement* course. For instance, the need for more practice in the design of analytical and hybrid rubrics and the continuous reinforcement of written guidelines, as well as the magnitude of having GSs exposed to the theory alongside the practice of authentic evaluation and, above all, the enormous value of guiding GSs to become social agents, leaders both academic and professionally.

Taking the impact that having specialists on evaluation would mean for EFL teaching and learning in Costa Rica, TEEs and TR conclude with a respectful request for educational and university authorities: to endorse specialization initiatives in different areas of applied linguistics.

## References

Álvarez, G., Villalobos N, & Nandayure Valenzuela. (2008). *The challenge approach: an innovative teaching and learning pathway*. I Congreso Internacional de Lenguas Modernas. U.C.R., San José, Costa Rica.

Council of Europe. (2001). *Common European framework of reference for languages: learning, teaching, assessment*. Edinburgh, Cambridge: Cambridge University Press.

Farrell, Thomas. *Reflective teaching: the principles and practices*. FORUM Vol. 36 No 4, October–December 1998. Retrieved from <http://eca.state.gov/forum/vols/vol36/no4/p10.htm>

Hancock, Charles R. (1994). *Alternative assessment and second language study: what and why?* ED376695 ERIC Clearinghouse on Languages and Linguistics Washington DC. Retrieved January, 2010. Retrieve from [https://oncourse.iu.edu/access/content/user/mikuleck/Filemanager\\_Public\\_Files/EF\\_L\\_Assessment/Unit\\_3/Hancock\\_Alternative\\_assessment\\_\\_\\_SL\\_study.pdf](https://oncourse.iu.edu/access/content/user/mikuleck/Filemanager_Public_Files/EF_L_Assessment/Unit_3/Hancock_Alternative_assessment___SL_study.pdf)

*Informe 1er año de labores programa Costa Rica multilingüe*. (2009). Retrieved January 11, 2010 from <http://www.ticotimes.net/temporary/Annual-Report-for-Costa-Rica-Multilingue-Spanish.pdf>

Kirschner, Paul A. et al. (2006). *Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching*. Lawrence Erlbaum Associates, Inc.

*La Gaceta N.61. Digital, Diario Oficial de Costa Rica*. (2008). Decreto N. 34425-MEP-COMEX, 11 de marzo y publicado el 28 de marzo. S.J., C.R. Imprenta Nacional. Año CXXX.

Mertler, Craig A. (2001). *Designing scoring rubrics for your classroom: practical assessment, research & evaluation*. Bowling Green State University. Electronic journal ISSN 1531-7714. PAREonline.net. Retrieved December 18, 2009. PAREonline.net/getvn.asp?v=7&n=25

O'Malley, J. Michael & Lorraine Valdez Pierce. (1996). *Authentic assessment for English language learners: practical approaches for teachers*. United States: Addison – Wesley.

Vande Hey-Klefstad, Jill M. (2005). *Developing critical reflection for pre-service education students: strategies to improve pre-service students' professional performance*. UMI Number 3210021. University of St. Thomas. St. Paul, Minnesota. Retrieved January, 2010 from <http://www.hiceducation.org/2005%20Final%20Program.pdf>



## FOOTNOTES

---

<sup>1</sup> Decree N.34425-MEP-COMEX March 11<sup>th</sup>, 2008.

<sup>2</sup> A detailed account is provided in the document "*Evolving from One's Depth to Teaching and Research Expertise: Chronicle of a Specialization Proposal.*"

<sup>3</sup> ELCL (in Spanish) stands for the School of Literature and Language Sciences at UNA.

<sup>4</sup> A detailed explanation is provided in the document *In-Service Teachers' Metamorphological Process in Evaluating English as a Foreign Language in Costa Rica.*

<sup>5</sup> Critical pedagogy entered the educational arena in the 1970's through the work of Paulo Freire (Brazilian educator) who incorporated the use of dialogue to help individuals become masters of their own thinking (Vande Hey-Klefstad p. 26).

<sup>6</sup> Task included in class agenda, October 17<sup>th</sup>, 2009.

<sup>7</sup> This is a major principle of the *Challenge Approach* (CHA), formulated by Nandayure Valenzuela Arce and adopted and nurtured by Gustavo Álvarez Martínez and Nuria Villalobos Ulate, all three professors at ELCL-UNA.

<sup>8</sup> As stated in the rating scale that was designed by the specialization candidate to gather additional data aimed at corroborating what was observed in class.

<sup>9</sup> Task included in class agenda, Friday, August 7<sup>th</sup>, 2009.

<sup>10</sup> Task included in class agenda, Friday, August 28<sup>th</sup>, 2009.

<sup>11</sup> As stated by the professor of the course in the rating scale that was designed and implemented by the specialization candidate, in accordance with the principles applied in the course observed.

<sup>12</sup> Defined as "the process of information gained through innovation in relation to the teacher's existing schema for teaching" (Pennington, qtd. in Farrell 10).

