Título: Rethinking Critical Language Pedagogy: Using Rubrics to test Oral Production of EFL Learners in oral Expression Courses

Autor 1 : Roberto Arguedas Zuñiga
Institución: Universidad Nacional de Costa Rica, Campus Coto
País: Costa Rica
Correo: Roberto.arguedas.zuniga@una.ac.cr
Correo:roberto.arguedas.zuñiga@una.cr
Autor 2: Cecilia Lopez Morales
Institución: Universidad Nacional de Costa Rica, Campus Coto
País: Costa Rica
Correo: cecilia.lopez.morales@una.cr
Autor 3: Andrés Ariel Robles-Barrantes
Institución: Universidad Nacional de Costa Rica, Campus Coto
País:Costa Rica
Correo: andres.robles.barrantes@una.cr

Resumen

En el aprendizaje y la enseñanza de un segundo idioma, los logros están usualmente relacionados con números o categorías determinadas por programas institucionales que son aplicados por los profesores de idiomas. Estas prácticas generalmente suelen sectorizar las personas en categorías de forma opresiva y direccionada. Para entender este fenómeno, es necesario primero analizar que la evaluación no es solo la aplicación de pruebas y el establecimiento de notas, comprendiendo que existe la posibilidad de generar un proceso pedagógico en el que, a partir de una valoración y evaluación inclusiva y colectiva, las diferentes personas que forman parte del proceso pueden contribuir en el proceso de aprendizaje. Este estudio pretende debatir y redefinir el concepto de evaluación para que sea replanteado en la comunidad académica como algo más que un enfoque positivista. Se encuentra enfocado metodológicamente en la pedagogía crítica y el análisis de los programas de los cursos de expresión oral de la carrera Enseñanza del Inglés.

Palabras clave: Pedagogía crítica, exámenes, evaluación compartida, evaluación

Abstract

In the foreign language teaching and learning process, the word achievement is usually connected to numbers or categories that are determined by institutional programs and applied by language instructors. The evaluation experience that includes appraising,

assessing, and testing in route to scoring has placed students into three cruel levels of knowledge and performance which are achievers, overachievers and underachievers. The first aspect that needs to be analyzed is the fact that evaluation is not just testing and scoring or just the mere application of tests to get a number but a process to obtain information to make decisions by applying the principles of shared assessment. This study is aimed at reframing and realigning the conceptualization of evaluation in order to influence the academic community positively towards a process that in education has been linked to a behaviorist approach in terms of positive and negative reinforcement but inclined towards punishment. The proposal is based on a critical pedagogical analysis of the syllabi of the oral expression courses using shared assessment to rethink the evaluation standards into a classroom strategy available to all the participants of the second language teaching and learning process.

Keywords: critical pedagogy, testing, shared assessment, oral expression courses, evaluation

I.   INTRODUCTION

Evaluation is a process that includes continuous observation and methodological guidance for the benefit of students. The vision and self-analysis of the teaching and learning process define the pedagogical path every language instructor should follow to achieve positive results during the last stage, which is scoring performance and competence to have a substantial result. However, there is a concept that needs to be culturally reconstructed to talk about effective evaluation in education and that is testing. As stated by Porras et al. (2012) "Language instructors need to undergo the route of introspection, where they can reflect and make pertinent decisions about their role as teachers and evaluators" (p. 2). It means that as active members of the process, testers in general need to understand and determine what testing is to plan, design, and conduct a liable process. Liability is supported by experience and the construction of the conceptualization by the roles of the testers and the test takers within a whole evaluation process grounded on shared assessment.

Testing is indeed a professional challenge stated as a responsibility when the interpretation of the results is linked to the purpose for what it was intended to serve. Though testing can also be employed for research purposes, this article will focus on two major uses in the educational setting: making decisions in terms of evaluation and being used as an indicator of proficiency. Notwithstanding, within the language teaching and acquisition field, testing has been focused on its second major use as an indicator of proficiency, that is, a way to measure abilities related to scales of achievement. Actually, in those two major uses of tests, the helpfulness and quality of the information provided by the results is upheld with different assumptions and considerations which are valid in the educational context. As professors, in our processes, a clear management of the different types of language tests

should predominate, as well as their role in evaluation and their suitability to be able to select and apply the most appropriate ones. Another important aspect to mention is that the interpretation of test outcomes in research is of both theoretical and applied concern to understand the nature of language proficiency, and to use that understanding into the second language teaching and acquisition process. As Bachman (1995) stated, the language tests can be classified according to five distinctive features: the purpose, or use, for which they are intended; the content upon which they are based; the frame of reference within which their results are to be interpreted; the way in which they are scored, and the specific technique or method they employ.

The second language teaching and learning process should be conducted through an evaluation experience supported by the conceptualization of fairness involving all its participants. The progress of each of the students must be reinforced by a concept that is not just part of education but education itself: social justice. Taking into account that the learning of a second language and its evaluation require the study of both instructional and functional dimensions of language, the process needs to be pedagogically transparent. There is an ethical link between how to teach a second language, how to learn that language, and how to evaluate that learning process. Crawford-Lange (1981) mentioned, in a Freirean view, that the teacher has to participate "as a learner among learners….[who] contributes his/her ideas, experiences, opinions, and perceptions to the dialogical process…[and] becomes one with the students" (p. 266) within the teaching and learning process while "the student possesses the right to and power of decision making" (p. 266); it is this right which constitutes the foundation of shared assessment as being meaningful and significant when evaluating language proficiency through critical language pedagogy.

II. METHODOLOGY

This paper is based on a document analysis where the researchers consider different pedagogical aspects to contrast and compare information in the curricular programs of the English Teaching major at Universidad Nacional, Coto Campus. Though the methodology does not require the administration of instruments to a specific group of participants, this method contributes to a qualitative analysis where it is possible to perceive and identify phenomena that influence the learning process. Producing a rich description of a specific event can contribute significatively to understanding the teaching-learning process. Regarding this methodological approach, Bowen (2009) emphasizes that "document analysis is a systematic procedure for reviewing or evaluating documents—both printed and electronic (computer-based and Internet-transmitted) material" (p. 27). In this case, the researchers analyze the syllabi of ten courses: Oral Expression: Society and Humanism (OE S&H 2017,2018 & 2019), Oral Expression: Commerce and Economy (OE C&E 2017 & 2019), Oral Expression: Science and Technology (OE S&T 2017, 2018 & 2019), and

Advanced Oral Expression (AOE 2017 & 2019). Each course expects learners to reach a level of proficiency which is divided in the following manner: B1 (OE S&H), B2 (OE C&E and OE S&T), C1 (AOE). Also, the methodologies of the courses are based on collaborative learning (CL) and inquiry-based learning (IBL). The oral performance criteria found in the syllabi rubrics are varied; thus, for the analysis, the researchers focus on different elements related to the course methodology, evaluation, and oral performance criteria. This research process is divided into four different stages: selection of the data, synthesis of the data, analysis of the data, and conclusions. Researchers follow an analytic path characterized by a process that "entails finding, selecting, appraising (making sense of), and synthesising data contained in documents" (Labuschagne, 2003, as cited in Bowen, 2009, p. 28).

Description of the Stages:

Table 1
Methodological Research Process

Methodological Stage	Description
Selection of the data:	Researchers consider the programs of the oral expression courses of the English Teaching major for this analysis. These are four different courses taken during the second, third, and fourth level of the major.
Synthesis of the data:	The information is arranged in a comparison table designed by the researchers. This table considers elements such as the students' expected level, course methodology, evaluation, and oral performance criteria which are the most relevant aspects that the analysis would consider.
Analysis of the data:	By considering the information set in the comparison table, researchers analyse all the elements presented. In this way, it would be possible to describe and understand some of the methodological aspects implied in the evaluation of the oral performance criteria in the selected syllabi.
Conclusions:	Conclusions are presented as a result of the research process and as an invitation for further and deeper studies in this pedagogical phenomenon.

Note: Designed by the researchers.

Researchers' Reflexivity: The researchers are three professors of the major whose teaching experience includes the oral expression courses. Their influences to carry out this research are grounded to the purpose of improving their pedagogical praxis and contribute to the debate regarding EFL evaluation for oral performance. As an ethical issue, it is relevant to consider that this research is a preliminar contribution for the analysis of the criteria, it would be necessary further and deeper investigation to answer all the questions that this proposal may generate since documentary analysis does not allow the researchers

to consider the perceptions of all the participants involved in the evaluation process. This is a method to identify and describe a perspective of the reality, not the reality itself.

III. RESULTS AND ANALYSIS

Methodology and evaluation of the Syllabi

In Table 2, the different types of assessment and evaluation in the oral expression courses are outlined. It is important to note that the only recurring evaluation is the board-based speaking test (BBT) in the English Teaching Major, for which Sevilla-Morales and Chaves-Fernández (2020) reported on the strong washback effects of this type of test, from the students' perspective. Though they recognized the lack of addressing professors' perspective, they also warned about the validity of such methods, as "seemingly analogous opinions in terms of the consistency between the BBT and the professors' methodological choices...violat[ing] the principle of transparency in language assessment…[along] with reports that some professors never update their teaching methods" (p. 233). Thus, upon examining the syllabi of the oral expression courses that employ this evaluation under a CL and IBL framework, the types of evaluation endorsed by professors in these courses and the criteria with which they test students' language proficiency and evaluate oral performance must be considered.

Table 2

Types of Evaluation Used in the Oral Expression Courses per Year

Evaluation in the Syllabi	2017			
	OE S&H	OE C&E	OE S&T	AOE
Impromptu Speeches	X	X	X	X
Oral Presentation	X	X	X	-
In-class Participation (Log)	X	X	X	X
Diverse group activities*	X	-	X	-
Midterm (pair dialogues)	X	X	-	-
Board-based speaking test	X	X	X	X

Evaluation in the Syllabi	2018			
	OE S&H	OE S&T	OE S&T	AOE
Impromptu Speeches	X	X	*	*
In-class Participation (Log)	X	-	*	*
Diverse group activities*	X	-	*	*
Midterm (pair dialogues)	-	X	*	*
Board-based speaking test	X	X	*	*

Evaluation in the Syllabi	2019			
	OE S&H	OE C&E	OE S&T	AOE
Impromptu Speeches	X	-	-	-
Oral Presentation	X	X	X	-
In-class Participation (Log)	X	X	-	X
Diverse group activities*	X	-	-	X
Board-based speaking test	X	X	X	X

Note. **Created by the researchers. X= included in the syllabus; - = not included; *= syllabus not available or retrievable**

A close second and third contender in most used evaluation are oral presentations and impromptu speeches. Oral presentations require students to research before individually presenting before the class, and only one syllabus allows for pair presentations, that is, there is a consistent belief that speaking must be evaluated more so individually than collaboratively; to exemplify, two courses (OE S&H 2017 and OE C&E 2017) propose a news article presentation. Additionally, only two syllabi typify oral presentations specific to

course contents: a business start-up presentation (OE C&E 2019) and a science exhibition project (OE S&T 2019), which also evidences a tendency to replicate the evaluation in syllabi through time and, therefore, suggests a lack of innovation and contextualization in presenting a generic description of evaluation. In-class participation seems to be a major concern in the oral courses as well, although, as Freire and Shor (2014) have mentioned, dialogue cannot be coerced and students have a certain right to remain silent; that is, the end goal of dialogue is not for everyone in the class to speak even when they have nothing to offer. Silence sometimes means that students are analyzing. Diverse group activities are also employed which include a myriad of activities, such as, but not limited to, the following: symposium, debates, discussion panel, forums, workshops recordings, tutorials, reports, news program, talk show, reality show, roleplays, monologues, dialogues. Lastly, the midterm is a speaking test that mimics the BBT or, in three specific cases (OE S&H 2017, OE C&E 2017, and OE S&T 2018) calls for paired dialogues.

Evaluation and Criteria

In comparing the evaluation of the syllabi and the rubrics (analytical and holistic) with which students' speaking performance will be assessed, it is important to delimit which criteria are the most relevant and how they are described. As the criteria may be isolated or combined (e.g., pronunciation and fluency) they were separated in order to make the analysis, as the descriptors referred to both. Though ten courses were considered, one did not contain rubrics; hence, it was impossible to analyze the rubrics and the underlying oral performance criteria. To begin, all the oral expression courses consider at least the following linguistic criteria: grammar, pronunciation, fluency and vocabulary. These first criteria require a special mention as they appear regardless of the professor designing the syllabus.

Six of the nine syllabi assess content and preparation; five of the nine syllabi included organization, task fulfillment, interpretation, analysis, synthesis, and involvement; four of the nine syllabi included comprehensibility, command, introduction, non-verbal strategic competence and conclusion; three of the nine syllabi included elocution and participation; two of the nine syllabi included communication skills and glossary; the least cited criteria, included in only one syllabus, included audiovisual (and other resources) aids, delivery, relevance, time, and topic and activity development.

In order to determine the participation of teachers in the elaboration of the course syllabi, they will be referred to as Prof A, Prof B, Prof C, Prof D and Prof E. That is, five professors designed the syllabi, and, as such, it is important to determine if some participated more than others and analyze the relationship between the criteria chosen for each course. Only three professors taught one oral expression course: Prof A (OE S&H 2017), Prof C (OE S&T 2017), and Prof E (OE S&T 2018). Nonetheless, of the least cited criteria only that of topic and activity development stems from the OE S&T 2017 syllabus

designed by Prof C. In fact, the other least cited criteria (one mention and two mentions only) are included by Prof B, who designed the syllabi for four different courses (OE C&E 2017, OE C&E 2019, OE S&T 2019, AOE 2019) and appear in the same course (OC C&E 2017 and 2018) which suggests recycling rubrics and evaluation criteria, raising the questions which Sevilla-Morales and Chaves-Fernández (2020) posited regarding decision-making, innovation and training on part of the professors, as well as tailoring the syllabus to different student needs and proficiency levels. A similar case occurs with Prof D, who designed three different syllabi (AOE 2017, OE S&H 2018, and OE S&H 2019) where the same criteria can be found.

Finally, there are many concerns regarding the descriptors of the performance criteria in syllabi rubrics and raises the question as to whether a collegiate or unified view of the criteria should be in place, as the same criteria will subjectively render different evaluations for students. For instance, grammar is associated with using simple and complex structures but also linked to a language learner's autonomy, which occurs with pronunciation as well, and is measured as something that is "shown" by the learner. In the same vein, "mastery" of content is assessed, assuming an idealized mastery can be achieved in any kind of content, and comprehensibility is viewed as "entirely comprehensible to native speakers of English." In short, whether a general consensus can be reached regarding the understanding of oral performance criteria is a matter that emerges through dialogue between all of the parts involved: faculty language professors, professors and students, and a coherence between teaching methodology, syllabus and curriculum design (stakeholders).

III. RESULTS OF THE ANALYSIS

The teaching process of a foreign language in terms of evaluation should be considered to be another classroom strategy to reinforce the learning experience where evaluation is understood as the most effective way to enrich the learning process and conduct changes to obtain reliable results. Testing in isolation should not even be understood as a primary device to define promotion. Nonetheless, "whether a test focuses on aspects of second language competence or performance, its construct validity is the overriding concern in its development and validation" (Roever, 2001, p. 86). In terms of the syllabi, this implies that the methodology must be coherently linked to evaluation and the oral performance criteria, which students must understand and help construct.

a) Testing and success

Testing provides outcomes to interpret if students are able to reach an apparent level of language proficiency and fulfill instructional objectives. However, it is not a methodological determinant to define students' knowledge, competence or performance. There are external and internal aspects that might affect the fairness of the process, as part of continuous

evaluation. Teachers' negative attitudes while testing, the use of inadequate pedagogical activities, the test's presentation, the lack of coherence between objectives and contents, the excessive use of memorization for oral performances and the lack of a consistent evaluation process prior to the testing stage might mislead students' answers and, therefore, the results. To this end, after presenting a critical discourse analysis of students' self-perception of language proficiency, Fallas-Escobar (2017) contended that "Diversified proficiency models used in class, coupled with first-hand experience interacting with them, should help students stop engaging in self-derogation of their current and evolving language proficiency and start reformulating the proficiency standards that they aspire to" (p. 281), in this case, for testing and success to make sense.

b) Testing assumptions

Evaluation takes over the most important role embracing its two main components: information and decisions. To justify those decisions, there are three relevant assumptions to take into consideration when gathering information through the use of tests. The first assumption links effectiveness and accountability. Professors must assume that the information about educational results is essential for effective formal education. Bachman and Savignon (1986) equated accountability to the degree of discharging responsibility and that "without accountability in language teaching, students can pass several semesters of language courses with high grades and still be unable to use the language for reading or for conversing with speakers of that language" (p. 380).

The second assumption bases the changes to the program and the improvement of the learning process by examining feedback (Bachman, 1995). In simple terms, feedback is "information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding" (Hattie & Timperley, 2007, p. 81). This is important because the information is of interest to both parties: students and teachers (Bachman, 1995), as they shed light on major inquiries. In this sense, feedback from a test is as effective from the evaluator as it is from the student, or "the test's most important stakeholder: the candidate" (Ryan, 2004, p. 20), bearing in mind that "the recognition that educational theory must be grounded in the interpretations of teachers, is not in itself sufficient" (Carr & Kemmis, 1986, p. 129). This reaction may be even more relevant, as it provides a continuum of action, given that feedback can be seen as natural "'consequence' of performance" (Hattie & Timperley, 2007, p. 81) and the decisions need not only be made from the teacher's perspective. The reality is, in turn, that the 'objectivity' behind scoring procedures of the test is questionable. Thus, it is as Hattie and Timperley (2007) suggest, that the role of feedback may be cyclical and multidirectional or "correspond to notions of feed up, feed back, and feed forward" (p. 86) to consider reactions to the feedback from the test takers as well, to engage in dialogue, promote shared assessment or a "type of

assessment that is more inclusive and democratic in nature than the traditional, authoritarian type of model" (Ryan, 2004, p. 19).

Finally, the third assumption states that the educational outcomes of the given program are measurable, and it is debatable. Most of the time, outcomes are used to provide information to clarify to what extent students' performance may be evaluated as competent and used to propose changes. Consequently, the qualitative information gathered from testing needs to be interpreted from two different positions: any outcome that is learnable is measurable, or no valuable educational objective can be adequately measured. Therefore, this study is aimed at critically analyzing the programs of the oral expression courses of the major in terms of coherence between methodology and evaluation.

c) Underlying EFL Testing Assumptions at UNA

To conduct evaluation and testing in a second language teaching and learning process, it is relevant to highlight what the program states as teaching. The English Teaching Major Program (Universidad Nacional, 2013) claims that the teaching process that prevails does not make learners become recipients of memorized information as knowledge but promotes new schemes based on logical, critical and real action, which opens new perspectives of the world. It also explains that teaching facilitates the construction and coordination of new strategies and thinking skills related to life, science, culture and or as a member of the academic personnel. Therefore, when testing, teachers should focus on questions and evaluative strategies to obtain results that contribute to that conceptualization of the teaching being pursued.

d) Testing subjectivity

The assumptions and the testing process could be highly arguable depending on educational programs: therefore, results are evaluated and used as qualitative information. Some considerations about tests to support and validate their results are the amount of testing, the quality of the information those tests provide, the reliability and validity of that information and thus the possible costs associated with the errors because those results will deal with decisions about teachers and programs. Within the testing process, achieving objectivity or understanding the role of subjectivity is determined by the educational expectations. Bachman (1995) explained the following:

> In an objective test, the correctness of the test taker's response is determined entirely by predetermined criteria so that no judgment is required on the part of scorers. In a subjective test, on the other hand, the scorer[s] must make a judgment about the correctness of the response based on [their] subjective interpretation of the scoring criteria. (p. 76)

When analyzing the implications of testing, there is a meaningful debate regarding the subjectivity implied in the different procedures undertaken for its purpose. Bachman

(1995) pointed out that besides scoring procedures in "objective" tests, "all other aspects of tests involve subjective decisions" (p. 76). These relations are based on subjective perceptions that are deeply linked to sociological behaviors and contributions to the learning process. Kemmis (2006), describing Foucault's philosophy on this matter, mentions that:

> Social practices like educational evaluation are not *sui generis.* Foucault teaches us that all these things are humanly, socially, historically constructed, and that they are social sites in which the practised eye can identify traces of the broader, deeper, older webs of social coordination…. (p. 36)

Furthermore, when students perform, they have to decide or ponder about the "correct" answer to a question, the appropriate way to solve an exercise or even "behave." In the case of language tests, oral presentations and compositions are subjectively graded even with the application of a rubric or scale (Bachman, 1995). Thus, an analysis in terms of achieving objectivity in testing is mandatory. However, it is necessary to question if subjectivity is not implied in the ideological understanding of that objective form of evaluating. Karier (1974) counter-argued that any evaluation is ideological:

> It is my understanding that evaluation is a complex process of assigning values to phenomena, while ideology represents that set of values and attitudes which go to make up the composite picture of the social and individual philosophy by which [people] in a given culture profess to live. In this context evaluation inevitably occurs within some kind of value orientation as part of an ideological framework. (p. 279, as cited in Kemmis, 1993. p. 41)

Every evaluation process has an ideological orientation; therefore, all tests are designed with a specific purpose [attempting] to fulfill an educational policy, an institutional parameter, or a systematic objective. When educators define the options working as distractors, the correct answers, and the structure of the questions, there is a subjective process being carried out. Hence, objective testing becomes a mirage that may hide what is really happening, or as Carr and Kemmis (1986) stated,

> The all-pervading influence of positivism has resulted in a widespread growth of instrumental rationality and a tendency to see all practical problems as technical issues. This has created the illusion of an 'objective reality' over which the individual has no control, and, hence, to a decline in the capacity of individuals to reflect upon their own situations and change them through their own actions. (p. 130)

Both in the language teaching setting and second language acquisition field, test developers as well as their users have aimed to concentrate and work on selecting the most appropriate tests according to the program and their intentions. Consequently, the types of tests and their suitability is based on understanding their uses, the information they provide, and how they contribute to establish a scientific support for changes, which could even be

seen as disenfranchising teachers. To this end, this positivist approach to testing has sought to provide a basis to objectivize facts or courses of action, which may be rigorously referred to as 'truth' or loosely defined as 'information.' By doing so, historically, "Scientific results [have] merely distinguished more effective courses of action from less effective ones and explained how outcomes occurred—not whether or not they should be allowed to occur" (Carr & Kemmis, 1986, p. 132) and this gives rise to an objection of rationality focused solely on " a conformity to the rules of scientific thinking, and, as such, deprived of all creative, critical and evaluative powers" (p. 133) by altering the process of theory or practice, not theory *and* practice, and relegating the teacher to a technical process, a complementary yet necessary process which cannot be viewed independently in the scope of critical theory.

IV. CONCLUSIONS

There is still much needed work in understanding testing as part of the evaluation process for foreign language learning (or the transition to ELF) not only at the standardized level but at the national, regional, and local level as well. Needless to say, the foreign language classroom is not exempt. Emerging pedagogy and methodologies surrounding English Language Teaching (ELT) have been questioned in terms of the overall effectiveness and suitability of testing standards around the globe, for "many current language assessments may largely ignore these facts and native standard is still the norm on the majority of national and international tests" (Fang, 2017, p. 58) and "we [professors] need to tackle the issue of how traditional approaches to English testing can be adapted to the situated language use" (p. 64)

Establishing constructs and learning outcomes for foreign language learners and how their performance will be evaluated is essential work of language professors and language programs to move away from an EFL oriented approach, or the "traditional accumulative means of testing, in which actual language performance cannot be tested... in order to, for example, focus on students' communication strategies, and how they use the language to fulfil different tasks in both academic and non-academic settings" (p. 65). In turn, emphasis should be placed on the sociocultural constructs to be evaluated, and the ties of language and identity, especially as acts of identity or a sedimented body of utterances and ritualized performances (Pennycook, 2006).

This, however, requires a deep reflection on part of stakeholders, education centers, language programs, professors and students alike in the decision-making process that shapes the testing paradigm. Pennycook (2010) admonished that "unless we constantly challenge our thinking about language, linguistics and applied linguistics, we run the danger of reproducing precisely those language ideologies we need to be opposing" (p. 9). In other words, the education community is solely responsible for endorsing the standards which they deem fit to teach and test, or otherwise, as Widin (2010) laments, "dismissing their own

expertise and indigenous knowledge, engaging in the practice of self-marginalization" (p. 60). It is imperative to emphasize the role of the language educator in possessing their own knowledge and feeling empowered enough to promote their own system or ideology. In the case of language learning, Pennycook (2010) reinforces this idea in that

> By making a notion of language as a local practice central to our activity we can overcome the tired debates about linguistics applied and applied linguistics, and the problematic notion of applying linguistic theory to contexts of practice. Once we grasp that language is a practice itself, we are no longer reliant on linguistic theory. (p. 9)

Especially without this contextualization, testing can often be perceived as detrimental to the teaching learning process. Shohamy (2005) affirmed that test takers believe that "tests are not at all indicative of their true knowledge…[and] that tests are detached from real-learning and from real-life performances" (p. 103) which is a call to action on the overall purpose and use of a test. She then proceeds to refer to testing as having detrimental effects and being used as disciplinary tools by citing that "they create winners and losers, successes and failures, rejections and acceptances" (p. 104), akin to Fallas-Escobar's (2017) findings. This requires questioning the use of tests, rubrics, their criteria and rubrics (Tierney & Simon, 2004) and "to critique the values and beliefs inherent in them" (Shohamy, 2005, p.108).

To accomplish this socially reflexive endeavor, it is also imperative to consolidate a notion of what communicative competence and performance implies. Notwithstanding, Pennycook (2006) immediately warned that "The notion of performance has been played down in language studies in several ways: from the inception of modern linguistics via Saussure's langue (system)/ parole (use) dichotomy, and more particularly in Chomsky's competence /performance division" (p. 58). He further added that "not only was the focus of linguistic enquiry on the abstract competence to produce grammatical utterances, but the means to arrive at such an analysis was not through everyday use of language (an empirical mess) so much as through rationalist introspection" (p. 58). This entailed an ideological implication that "performance of language in a social and cultural domain was not of interest" (p. 59). As such, the understanding, at least conceptually, of the underlying competencies has been expanded by several authors (Bachman, 1990; Canale, 1983; Canale and Swain, 1980; Pennycook, 1985) beyond the grammatical, including sociolinguistic, discursive, strategic, paralinguistic, organizational and pragmatic competencies (as cited in Pennycook, 2006). Though this may be seen as aggregate value, the competence/performance distinction has "focused attention on underlying competence as the driver of production, rather than seeing competence as the product of performance, or indeed doing away with the distinction completely" (p. 59). By examining such positions, one can immediately understand the intricacy underlying testing that encompasses the almost endless or infinite

contexts in which language can occur, or be performed, akin to the Chomskyan view of recursion, the infiniteness of human language production and creativity, but not limited to it; in other words "what ties performances together is not a competence that lies within each individual but a wide array of social, cultural and discursive forces" (Pennycook, 2006, p. 60) in iterable events that may be scripted or unscripted, rehearsed or unrehearsed. In these contexts, and in the ELF view, at least, "multilingual speakers will use English for utilitarian purposes with a pragmatic attitude" yet "they won't [necessarily] develop a cultural affinity with the language or attempt to represent their identities through English" (Canagarajah, 2006, p. 199).

In short, a revision at how tests are designed, what constructs underlie testing and what exactly can be perceived as communicative competence and performance is needed from a local perspective, one where teachers and learners' voices are heard and identified. In this way, constructs, models and frameworks as part of the testing process will always be defined by a transformative context that endorses empowered professors and students' own critical language pedagogy.

REFERENCES

Bachman, L. (1995). *Fundamental considerations in language testing*. New York: Oxford University Press.

Bachman, L. F., & Savignon, S. J. (1986). The evaluation of communicative language proficiency: A critique of the ACTFL oral interview. *The Modern Language Journal*, *70*(4), 380-390.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27.

Brown, H. D. (2004). *Language assessment principles and classroom practice*. NY: Pearson Education. Retrieved from http://images.pcmac.org/Uploads/JeffersonCountySchools/JeffersonCountySchools/Departments/DocumentsCategories/Documents/Language%20Assessment%20-%20Principles%20and%20Classroom%20Practices.pdf

Canagarajah, A. S. (2006). Negotiating the local in English as a lingua franca. Annual *Review of Applied Linguistics, 26, 197-218*. http://www.personal.psu.edu/asc16/pdf/ARALpaper.pdf

Carr, W. & Kemmis, S. (2003). *Becoming critical: education knowledge and action research*. Routledge. https://www.academia.edu/31301744/Becoming-critical

Crawford-Lange, L. M. (1981). Redirecting second language curricula: Paulo Freire's contribution. *Foreign Language Annals, 14*(4), 257–268. doi:10.1111/j.1944-9720.1981.tb01644.x

Fallas-Escobar, C. (2018). Diversifying proficiency models in EFL programs. *Revista de Lenguas Modernas*, *27*. https://doi.org/10.15517/rlm.v0i27.32146

Fang, F. G. (2017). English as a Lingua Franca: Implications for pedagogy and assessment. *TEFLIN Journal, 28*(1), 57-70. http://teflin.org/journal/index.php/journal/article/viewFile/418/279

Freire, P., & Shor, I. (2014). Miedo y osadía. *La cotidianeidad del docente que se arriesga a practicar una pedagogía transformadora.* Buenos Aires: Editorial Siglo XXI.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. http://www.columbia.edu/~mvp19/ETF/Feedback.pdf

Kemmis, S. (1993). Foucault, Habermas and evaluation. *Curriculum studies, 1*(1), 35-54. DOI: 10.1080/0965975930010103

Kurilovas, E. (2018). On data-driven decision-making for quality education. *Computers in Human Behavior, 107,* 1-9 . doi.org/10.1016/j.chb.2018.11.003

Pennycook, A. (2006). *Global Englishes and transcultural flows.* Routledge. https://doi.org/10.4324/9780203088807

Pennycook, A. (2010). Critical and alternative directions in applied linguistics. *Australian Review of Applied Linguistics, 33*(2), 16-1. https://doi.org/10.2104/aral1016

Porras, M., Cubillo, D., Álvarez, G., & Valenzuela, N. (2012). *Testing and measurement as a multipurpose mirror: the need to ignite introspection in language teachers*. Paper presented at the Congreso Internacional de Educación Superior, Universidad Nacional-West Chester University, UNA, Campus Omar Dengo, Heredia, Costa Rica. http://hdl.handle.net/11056/2887

Roever, C. (2001). Web-based language testing. *Language Learning & Technology, 5*(2), 84-94. http://llt.msu.edu/vol5num2/roever/default.html

Ryan, D. E. (2018). Consider the candidate: Using test-taker feedback to enhance quality and validity in language testing. *e-Teals: An e-journal of Teacher Education and Applied Language Studies*, *5,* 1-23. https://www.uv.mx/dcia/files/2018/01/13086.pdf

Sevilla Morales, H., & Chaves Fernández, L. (2020). Efectos colaterales de los exámenes orales por tribunal. *LETRAS, 68*, 199-238. https://doi.org/10.15359/rl.2-68.8

Shohamy, E. (2005). The power of tests over teachers: The power of teachers over tests. In D. J. Tedick (Ed.), *Second language teacher education: International perspectives* (101-111). Routledge.

Tierney, R., & Simon, M. (2004). What's still wrong with rubrics: focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research, and Evaluation, 9*(2), 1-7. https://doi.org/10.7275/jtvt-wg68

Universidad Nacional. (2013). *Plan de estudios de la carrera de: Bachillerato en la Enseñanza del Inglés.*

Widin, J. (2010). *Illegitimate practices: Global English language education. Clevedon, England*: Multilingual Matters.