

The Role of Assessment in the Language Classroom: An Action Research Report¹

AUDREY DON, UNIVERSIDAD LATINA DE AMÉRICA, MORELIA, MICHOACÁN

Testing is and always has been an issue in language teaching. There are times we feel we spend too much time on it or we feel the tests don't really reflect what students have learned; we wonder if tests are too difficult / too easy; if they are appropriate to the age / language level; the issues are many and varied.

I first became interested in the area of testing and evaluation when I came across the following. The author mentions four important factors that teachers should be aware of in relation to testing:

1. Test *what* was taught.
2. Test in the *way* we taught in class.
3. Test what learners *can do* (not just focusing on what they can't / don't do well)
4. Allow for the *creative use* of language. (Shrum & Gisan, 2000, p. 292)

The first point I felt was *not* a major concern in my case. However, I began to think about the *way* I test and the *way* I teach. This was the first thing that made me think there was something incongruent.

First of all, I thought about all the kinds of activities I engage my students in during regular class time (many varied activities immediately rushed to mind; the list was long). I *most* definitely teach in many different ways, using a variety of techniques and practices frequently, yet I tend to test in a fairly rigid pattern with the same very limited factors taken into consideration each time: a written test completed individually by students--in silence--that includes vocabulary, grammar-in-context exercises, reading comprehension, dialogue completion, etc. as well as a listening component. Although the written tests I administer do require students to use the language in different contexts (my problem is not with discrete items as are found in traditional tests) the problem is more with *the way* students are tested. In this paper, when I refer to testing, I will be referring specifically to the tools I use to evaluate my students. I will describe the kinds of tests that I have been using as well as the other factors that I take into account to arrive at the final grade for students during the semester.

If I were to conjure up an image of my students during class time I thought of my students sitting in pairs or in groups, working collaboratively. Yet when I thought about the class during testing, only one picture came to mind: my students sitting in rows, heads down, writing. The two images did not coincide.

To begin with, I looked at the way students worked in my classes. In a typical day in our classroom, students frequently work in pairs or in groups on a myriad of different tasks from songs to conversations; games and puzzles, TPR activities, and

¹ This is a refereed article.

interactive tasks that require students to use the language for some communicative, meaningful purpose. Yet, when I examined my grading breakdown, I realized that the major form of assessment and most of the semester grade was dedicated to one major format: a written test.

For years I have thought that my grading system (my assessment procedures) was effective, fair and representative of a communicative syllabus. On closer examination, I realized that 60% to 80% of the assessment for the semester was dedicated to written tests: from the short weekly quizzes (20%), the longer unit or two-unit partial examinations (20%) to the final semester written test (20%). I also included listening comprehension tests and quizzes (20%) and usually an oral grade based on one or two oral proficiency interviews toward the end of the semester (20%).

I also considered that my classes allowed for the *creative use* of language, but questioned whether my testing practices did also. I was not convinced. When I corrected the tests, what did I focus on by marking things wrong? I drew student's attention to what they had done *wrong*, not what they had done *right*; what they *couldn't do* rather than on what they *could do*. I realized that as far as my *own* teaching was concerned, I was failing to attend to items 2, 3 and 4 from the list above (in terms of my testing practices). Here is where I began my investigation on the topic.

Jim Cummins writes that "Assessment and instruction are two interlocking and interdependent components of any educational program" (O'Malley, Valdez Pierce, 1996, p. iv.). It seemed to me that I had not managed to develop the required interdependency between the two.

Often, in fact, testing in my classes is an add-on—almost like time-out from instruction "... to attend to the business of getting a grade for the student's report card" (O'Malley, Valdez Pierce, 1996, p. iv.).

I had to go back to the basics: Why do we test / evaluate? Clarke & Agne indicate there are several purposes for assessment that can be grouped into four main areas:

1. To focus student learning (to inform and guide students)
2. To focus on teaching (to inform on day-to-day teaching and to evaluate the effects of teaching)
3. To improve systems (to evaluate systems, i. e., the curriculum).
4. To influence policy and planning (to evaluate programs / to inform the public) (Clarke & Agne, 1997, p. 296)

Of these, we as classroom teachers are primarily concerned with the first two points.

Classroom testing procedures, according to Cohen, have undergone some improvements in recent years. He talks about "...a shift from using assessment as a way to keep students in their place to using assessment as a way to help students find their place in a school and the world community of language users" (Cohen, 1994, p. 3). He also draws our attention to research carried out by Sho-

hamy in 1985 (Cohen, 1994, p. 4, based on Shohamy, 1985). He lists both some questionable classroom practices as well as some promising classroom practices that have served to further define for me what we should and should not be doing with testing:

What we should avoid:

- Administering tests instead of giving instruction.
- Using tests as the exclusive measure for grading.
- Returning tests to students without offering corrections or explanations.
- Using only one testing method.
- Giving tests students did not know how to take.
- Taking too long in returning tests.

What we should be aiming for:

- Broadening the scope of what is included in assessment from tests alone to a variety of formal and informal assessment techniques.
- Viewing assessment as an opportunity for meaningful interaction between teacher and students.
- Judging students on the basis of knowledge they have, rather than on what they do not know.
- Using assessment measures intended to help learners to improve their skills.
- Training the test takers in test-taking strategies if performance on the assessment task could benefit from such training.
- Returning the evaluated tests promptly.

(Cohen, 1994, p. 4, based on Shohamy, 1985)

Further research matters have caused me to question the way I evaluate my students, particularly the use of the more traditional written test format. According to Ellis, Larson-Freeman and Long (Cohen, p.16) "rarely is there a one to one correspondence between what is taught and what is learned." If we are to consider this, we might wonder why, on completing a unit, topic or theme, we immediately proceed to test our students. We also have to keep in mind that not all students learn the same things in the same ways, nor do they learn at the same rate. Yet we expect all of our students to perform equally well on our written tests. If, as Nunan (1999, Ch. 4) points out, learners learn many things simultaneously and imperfectly, our expectations of student performance on written tests may be quite unrealistic.

Other interesting studies carried out in the 1970s and 1980s (Lightbown and Barkman 1978; Lightbown 1983; Pica 1983; Ellis 1984, to mention a few) found evidence that language instruction does not alter the pattern of grammar acquisition (Larsen-Freeman & Long, 1991, pp. 304-307). I was fascinated to read that students learn certain verb morphemes in the same order, whether they have had formal classroom instruction or have learned outside the classroom: First, -ing; second, regular past tense; third, irregular past tense and fourth, third person present -s.

This comes to me as quite a surprise given that most EFL textbooks I have come across expect students to learn the simple present (including third person singular -s) before any of the other verb forms. (Since reading this, I have not penalized my students at the beginner level on tests for forgetting the "s" on the third person present).

Another point worthy of mention is the importance of attending to student autonomy not only in teaching practices but in testing and assessment practices as well. If we are to encourage students to become more independent learners, it is essential that assessment become a tool for them as well as for us as teachers.

From what I have read, it appears that assessment, like teaching which has over the years moved away from favoring any one methodological focus to favoring an eclectic approach, should include a variety of methods. What I also hoped to explore in my Action Research Project were some alternative ways of testing / assessing my students that reflect an interactive language class.

Below is a summary of some of the questions that I needed to consider in regard to the initial problem I posed for this project on testing and assessment:

- What measures can I take to bring testing more in line with what I do in the classroom in terms of the way I teach? How can I develop the interdependency between teaching and testing?
- Aside from the traditional written test, how can I broaden the scope of the assessment I am using?
- What other *formal* measures could I use? What *informal* measures could I use?
- How can I better use testing as a tool for students to measure their progress and guide them to improvement?
- How do students see testing? What purpose does it serve for them?

Below is also the summary of the conclusions I have collected from various sources and the points that I was most interested in for this project:

1. Integrate testing and teaching.
2. Sample student performance in daily learning activities: informal assessment (TALK scores: T= Talking, A= Accuracy, L= Listening K= Kindness and cooperation (Shrum & Glisan, 2000, Appendix B.2, pp. 392-394), dictations, collaborative oral exercises, etc.
3. Test in a context that is meaningful for students.
4. Use formative ² (assessment without grades that gives students and teachers the chance to repair) as well as summative grades (grades assigned to measure learners' advances—value added).
5. Use multiple means of assessment.

² Although I included some forms of assessment that could be regarded as formative, for the purposes of this Action Research Project I focused only on summative assessment for grading purposes.

6. Use not only teacher assessments but peer and student self-assessments as well.

My class

My class was a first year basic level EFL class I had at a small private university in Morelia, Michoacan, Mexico that met for four hours per week. The age of the students ranged from 18 to 23; all of the students had studied English for several years at primary and high school level but had been unsuccessful due to either the large number of students in the class, poor learning conditions (teachers spoke to them in Spanish rather than using the target language) or low motivation. They had all experienced the sensation that English was difficult for them, yet they were all eager to be able to speak in English as this had been given little attention in their past experiences, which had been more focused on grammar exercises, filling out workbooks, etc.

Plan of Intervention

My research provided me with a very good source of alternatives for testing that I tried out over a period of two weeks at the end of a two-month period, approximately 32 hours of class time.

I compared the very limited traditional form of testing (the written test and the audio test) with a series of alternative assessment measures with the aim of bringing my assessment measures more in line with the kinds of activities we do on a daily basis in the classroom.

My aim was to create the interdependency that Cummins speaks of between classroom activity and assessment. I gave an initial questionnaire to my students that compared the kinds of activities we do on a daily basis compared to those we do for assessment (which looked at the traditional test format). I then gave them another questionnaire at the end of the two-week alternative assessment period to compare the two forms of assessment and to get opinions from students on their preferred form of assessment (based on the reasons they gave for their choice).

The following is a breakdown of how I would traditionally grade (first period) and how I graded in the alternative grading period (second period):

First two-month period (grading system and testing)

- 50 % 1 formal written test (based on the content of the textbook we use)
- 20 % 1 quiz (dictation of questions which were answered by students)
- 30 % 1 listening comprehension test (found in the Teacher's Manual of our textbook)

Second two-month period (Action Research and Alternative Testing)

- 10 % 1 listening comprehension test (as in the first period)
- 10 % 2 quizzes (dictations of questions that were then answered by students, as in the first period)
- 10 % 1 error correction quiz (students' correction of the classes' most common mistakes--those that had arisen from class work)

- 20 % 1 cooperative speaking test (pair work; peer- and self-evaluation) (idea adapted from Tim Murphy in Brown, 1994, see Appendix)
- 20 % TALK scores (oral testing of daily classroom speaking activities—concept and rubrics from Shrum & Glisan, 2000, Appendix 8.2, pp. 392-394)
- 20 % 1 cooperatively designed written test including items from classroom task-based activities (used for both TALK scores and inclusion on the written test and material covered in the textbook).
- 10 % 1 descriptive paragraph of a family member—done in-class as a test

Finally, after conferencing with students and taking some time to reflect on the results of the modified assessment plan, I drew some conclusions as to the impact this study would have on my classes in the future.

Analysis, Reflection and Conclusions

The questionnaires I gave my students after the first period of traditional assessment were enlightening: Clearly the majority of the activities we were involved in on a daily basis were *not* considered in the evaluation process. I asked students after the second period of alternative assessment to compare the two forms of evaluation in a questionnaire similar to the one that I had given them after the first period of evaluation. The objective was for them to check which activities (the same list of activities that we often did in our regular classes) we evaluated in the first period compared to the second period. The results confirmed that our assessment in the second period had included a greater number of the activities that we did regularly in class.

My students had calculated that we evaluated an average of about 30% of the regular classroom activities in the first period as compared to between 60-70% in the second period.

In the second period, we had managed to evaluate a much greater percentage of the activities that we did in class on a regular basis.

A very important moment during the second alternative assessment period I carried out for the Action Research Project was the self- and peer-evaluation I included in the summative evaluation on the Cooperative Speaking Test (Appendix).

The first part of the evaluation form asked students to reflect on how well they had been able to carry out the tasks. They then went on to the evaluation for grading purposes. Although this was a summative evaluation form in essence, it most certainly contemplated formative evaluation in that students had to reflect on their performance, as the first part of the evaluation was reflective and not related to their grade.

The final interviews I conducted with students were also very revealing. We spoke about the two evaluation periods before they filled out the final questionnaire. First of all I gave them the figure calculations of the two exam periods and explained again, to remind them, what items had been taken into consideration in the two periods. We then had a round table discussion in Spanish about the advantages and disadvantages of the two systems. Interestingly, to begin with, some students said straight away that they preferred the first period because there was

less work: They just came, did their tests and were finished. Then other students started to talk about how they had not done so well on the written test and that the second period had taken into consideration a wider selection of material and that it had been better for them to be evaluated constantly. One student commented that if you missed class on the day of the exam, you could miss a huge percentage of your grade—more food for thought. Another student offered the opinion that she had learned more in the second period because she knew we were having many different evaluations and had made an extra effort to review and be constantly up-to-date. As a result of the backwash effect, she was encouraged to work more consistently. I then left students to complete their questionnaires telling them there were *no right or wrong answers*: They were to honestly write down which form they preferred and I would only use their individual choice to evaluate them for the last period. If they chose the first period form of evaluation I would only count the final written exam, the listening exam and the oral interview. If they preferred the second period, I would carry on evaluating in multiple ways during the entire final evaluation period.

An interesting result of the questionnaire is that everyone chose the second period for a combination of the reasons we had discussed in the whole class discussion as well as new ideas that had not been discussed. Two students wrote that they felt the second period of evaluations had given them the chance to improve; another student elaborated on this by saying that in the second period if they did badly on the written test they could still make up their grades on the following evaluations. Another reported that he thought it was easier in the second period without the pressure of just one written test, while yet another reported that she had felt more involved in the subject matter.

So, I was not the only one who was more content with the way we had evaluated their work: We had all been involved in the assessment / evaluation process and the evaluations had in fact seemed more woven into the learning process and had not been only something additional.

The Tools

I will return briefly to the alternative evaluation tools I used and speak of the advantages I saw in using them.

I decided to include an assessment item that reflected the work we had been doing during the semester on error correction. Students had to keep a log of their own errors, of the situation in which they occurred, and their own remedy to help them remember the correction for the future. I have periodically shared these anonymously with the whole class and we have analyzed them as a short class activity. For the quiz, I gave them a selection of sentences with the typical errors that had come up in their class work and homework and they had to correct them. The students were familiar with this type of exercise and were very successful at it. They even performed well almost two weeks later on the section of the written test that dealt with physical descriptions of people (a parallel of the in-class activity).

The short descriptive paragraph they had to write was also similar to the kind of work we had been doing for the last month on describing people. I had them describe one member of their family and tell me about their interests, likes and dislikes. They had to bring a picture the next day so we could match the description with the picture (in one of our classes we had done a similar activity describing a famous person). The Interactive Speaking Test was an idea I adapted from Tim Murphy (Brown, 1994) that gave complete autonomy to students in much the same way as they would carry out pair work in our regular classes. I had never tried anything like this before and was somewhat hesitant of giving the students so much responsibility on a grading exercise, but could not have hoped for a better result. The students carried out the evaluation task as indicated and went about evaluating themselves and their peers fairly, critically and responsibly. It was definitely a vote in favor of student autonomy and has made me curious to look for and create other such opportunities for my students. All agreed that it was a fine learning experience.

The collaborative written test gave my students, for the first time, a role in preparing the written test. It was a valuable class activity--collaboratively recalling and registering from memory the content of the classes during the two-month period to be evaluated--as well as the actual exercise of designing test items to be included on the test. As this was the first time we had done such an activity, I was unable to use the test items they had designed; they were either inappropriate, inaccurate or invalid test questions. However, we did use their ideas as practice exercises and they provided some good input to clear up doubts and confusion before taking the test. So the entire exercise had been an extremely useful learning experience.

This semester was my first attempt with the TALK scores--I think it is a great new tool as it does not take time out from the class: Students are evaluated on their *Talking, Accuracy, Listening, and Kindness and cooperation* as they partake in any pair or group work activity in a regular class.

Conclusion

Many of the things I tried out for this Action Research Project were, in fact, very closely linked to my normal classroom practices, which was my major objective to start with. I have obtained many insights as a result of this project and am now convinced that we must use a greater variety of tools to assess our students. It is essential that we keep in mind that testing and evaluation should not be time out from the classroom, nor should it always be a separate activity, distinct in nature and style from a regular classroom activity. I am energized to continue exploring and experimenting with the new tools I have discovered. I decided not to go into the area of formative evaluation but this will now be my long-range objective. I am convinced that this experience was formative for both my students and for myself. We have all learned not to dread evaluation time and many students, as well as myself, witnessed how the constant process of evaluation kept them more focused and more aware of what they needed to do as language learners. It has

made them realize that the effort they put in produces results when one is constant. Unlike the usual cramming, last-minute studying for the next day's exam, the students learned the importance of on-going review and regular participation. Those who were constant felt the difference in their progress and I too witnessed their progress.

This project has helped me to integrate testing with what we do in the classroom on a regular basis. There must definitely be more of a fusion of regular classroom activities with evaluation activities. In fact, sometimes they were one and the same. I had thus managed to attend to Shrum and Glisan's second point: Test in *the way you teach*.

I felt that the activities I had chosen were definitely more meaningful for students and there was not the usual pre-exam anxiety when we carried out the evaluations. Students were serious about their work, and there was a healthy tension as they went about the tasks at hand, not the nervousness that produced the negative reactions I had seen so often in the past when it came time to take a test. I had also made space for Shrum & Glisan's third point: Test what learners *can do* (not just focusing on what they can't / don't do well).

I feel there was a good selection of varied activities, as there are in our normal classes. I had managed to achieve multiple means of assessment as well as allowing for the *creative use* of language while assessing (Shrum & Glisan's fourth point).

When I returned to the issue of grading and the weight given to each of the different aspects, I felt more assured that my grading policy was now more representative of the kind of work we carry out in class. I compared what I had traditionally evaluated and the respective percentages and made a projection of how I will probably assess in future. There are noticeable differences in the way I have distributed the weighting of each section now and I feel confident that it is an improvement on my old system. It is now more representative of classroom work and of the time we spend on each of the different skills. The traditional written test has not been disposed of, yet it no longer has the main role in the overall scheme of grading.

I will continue my search for ways to improve and modify in following semesters, although I feel I now have a solid base for more meaningful evaluation as a result of this study.

Finally, in addressing the issue of self- and peer-evaluations, I am sure I have found (as others before me have also) one more important key to changing students' attitudes to evaluation. If, as two of my students wrote on their final questionnaire, the evaluations we do in our language classes actually help students to improve, to become more focused in their studies, to encourage them to work consistently and to become more involved in the process of learning, then we have most definitely done our students a service as educators.

REFERENCES

- Brown, H. D. (1994). *Teaching by Principles: An Interactive Approach to Language Pedagogy*. Englewood Cliffs, New Jersey: Prentice Hall Regents.
- Clarke, I. and Agno, R. (1997). *Interdisciplinary High School Teaching: Strategies for Integrated Learning*. Boston, MA: Allyn and Bacon.
- Cohen, A. (1994). *Assessing Language Ability in the Classroom*. Boston, MA: Heinle & Heinle Publishers.
- Larsen Freeman, D. & Long, D. (1991). *An Introduction to Second Language Acquisition*. New York: Addison-Wesley Longman.
- Nunan, D. (1999). *Second Language Teaching and Learning*. Boston, MA: Heinle & Heinle Publishers.
- O'Malley, J.M and Valdez, Pierce, L. (1996). *Authentic Assessment for English Language Learners: Practical Approaches for Teachers*. New York: Addison-Wesley Publishing Co.
- O'Maggin-Hatley, A. (1993). *Teaching Language in Context*. Boston, MA: Heinle & Heinle Publishers.
- Shrum, J.L. & Gissoni, F.W. (2000). *Teacher's Handbook: Contextualized Language Instruction*. Boston, MA: Heinle & Heinle Publishers.

Appendix

Interactive Speaking Test: Student A

My name: _____

My partner's name: _____

***** Nota. No se puede usar el español en el examen. Si tu compañero usa el español nota UN PUNTO abajo por cada palabra en ESPAÑOL. Se deduce el total de puntos de la calificación final.*****

Registro de puntos por hablar en español: _____

STUDENT A:

1. Ask your partner 5 questions about his / her family. Write the information below:

- 1.
- 2.
- 3.
- 4.
- 5.

2. Ask your partner 5 questions about his/ her likes and dislikes. Write the answers below.

Use your partner's name. (Example: Audrey likes coffee.)

- 1.
- 2.
- 3.
- 4.
- 5.

Total 10 marks.

Minus number of words in Spanish.

Interactive Speaking Test: Student B

My name: _____

My partner's name: _____

***** Nota. No se puede usar el español en el examen. Si tu compañero usa el español nota UN PUNTO abajo por cada palabra en ESPAÑOL. Se deduce el total de puntos de la calificación final.*****

Registro de puntos por hablar en español: _____

1. Ask your partner 5 questions about favorite weekend activities. Write the answers below. Use your partner's name.

- 1.
- 2.
- 3.
- 4.
- 5.

2. Ask your partner about his/ her likes and dislikes (DON'T ask the SAME questions as your partner *****)

Write the answers below. Use your partner's name. (Example: Audrey likes coffee.)

- 1.
- 2.
- 3.
- 4.
- 5.