3. The Strange World of EFL Testing

Has psychology no place in testing?

Mario Rinvolucri, Pilgrims, UK

Abstract

Do you find the world of Alice through the Looking Glass unsettling, thought-provoking and deeply strange? This is precisely what I feel about the world of language testing, with its breath-taking disregard for the person of the test-taker.This article will give you a look at tests from the point of view of the test-taker rather than the usual point of view of one who must give the test. It will explain how the test situation is basically unfair for many reasons not the least of which is that the test-taker is forced to accept the examiners' frame of mind. The second part of the article deals with how a humanistic test could be constructed.

A look at a classic of Testing Literature

To illustrate what I mean, let us take a look at *Language Testing in Practice*, (Lyle F. Bachman and Adrian. S. Palmer, Oxford, 1996). This solid tome, a thorough work of well-researched seriousness, runs to 377 pages. Not more than 10 pages deal with the psychology of testing, that is to say the psychology of test-takers. The main statement that Bachman and Palmer make on candidate psychology comes on pages 114-115:

"As noted in Chapter 4, the test-takers' responses to the characteristics of the test environment and tasks can potentially inhibit or facilitate optimum performance."

The authors then list three aspects of testing that may affect some candidates' ability to acquit themselves well:

"**1.**...... Test-takers' familiarity with test setting may determine, in part, their affective responses to test tasks. When there is a high level of correspondence between the characteristics of the target language use setting and tasks on the one hand, and the test setting and tasks on the other, we may be able to assume that test takers will have a generally positive affective response to the tests and test tasks.......

2....... We would generally expect that test-takers who have the relevant topical

knowledge will have positive affective responses to the test and test tasks.......

3. Finally, test takers' general levels and profile of language ability can influence their affective responses. Test-takers who have high levels of language ability are likely to feel positive about taking a language test, while less proficient test takers may feel threatened by the test.

To summarise the authors' thoughts in simpler language: if a given test seems to be measuring language use they will need in real life, the candidates will feel happy; if they know the answers to the test questions, the candidates will feel happy; and, finally if their language level is high, test takers will be happy sitting tests.

Is that really all there is to say about affectivity in language testing?
Yet Bachman and Palmer are honourable men who have advised the UCLES examination board in Cambridge, UK and many other exam authorities. Their book is convincing when it comes to discussion of various types of validity, reliability, construct development, scoring criteria, scoring methods, scoring procedures and scoring scales, but what about the human being who sits at the centre of all these conceptualisations, the **person** taking the exam/test?

From my reading (limited) of the testing literature in EFL, little has been written about the student, the human being, invited or forced into the crisis situation of the exam room. Without consideration of the human factors in testing, what is the use of elaborating scientifically honed and perfected tests?

The built-in psychological unfairness of the test situation
Major exams have a different psychological effect on different individuals. In my own particular case, tests often filled me with a feeling of adrenalin pumping, joy at performance, a feeling of challenge and exhilarating risk. The effect they had on my brother was largely destructive: his writing hand trembled so much in his 16 plus UK State exams that he could hardly hold a pen. He passed only one sub- ject and this "failure" governed the path he has taken through life. According to a teacher who dealt with us both, Bernard was noticeably more intelligent than me. I would submit that the British State's academic judgement of these two brothers at 16 was grossly inaccurate because it put Bernard in a situation he could not bear while offering me an ideal circus ring to show off in. I jumped through the hoops

with more glee than awareness or dignity.

The EFL testing experts do not concern themselves with cases like Bernard's. Words like *anxiety, panic, fear, crisis, stress* do not figure much in the indexes of their books. You have to go to sources like *Journal of Behaviour Therapy and Experimental Psychiatry* (No.3, 1972). to find the work of people like T.K.Beck who produced video-taped scenes for desensitisation of test anxiety. The scenes on his video include these:

* a person tossing and turning the night before an exam is to be taken
* a typical classroom with pupils talking nervously before the instructor arrives. He comes in carrying the exam papers.
* a close up of time slipping by as the anxious student writes frantically on official paper.

Are manifestations of anxiety and stress in the face of exams rare occurrences that only affect that tiny minority of the student population who need psychiatric help, or are the scenes above typical of what quite a large number of exam takers live through? I have yet to find, in the literature, any comprehensive list of the ways that people cope with pre-exam stress but here are two idiosyncratic examples.

a) In mid teenage this highly successful professional woman did ballet exercises from 6.00 till 8.00 am on the day of the exam. She would thus go into the exam with a relaxed and slightly tired body and a very alert mind.

b) A man who now runs and markets a major language exam, used to smuggle an old pair of slippers into the exam room. He would sneak his feet into them and a sensation of comfort would come over him. With his stress levels thus lowered he reckoned he could write much better papers.

These two people managed to cope with test-generated tension creatively and successfully. Many people, like these two, manage to cope with the internal crisis situation that an exam can generate, but there may be a serious price to pay in terms of unhappiness. The words that follow are those of a Spanish EFL teacher

on a TT course at Pilgrims in the UK:

"Yesterday I was talking to some of my friends about university and student life, and most of us thought it was an experience we didn't want to go through again. All the pressure of exams and results was too hard to make us want to repeat it: one of us said that after finishing her studies she still had dreams about having to pass a test again, and not being able to do it." (Humanising Language Teaching, www.hltmag.co.uk *Year 4 Issue 1, Jan 2002, Readers' Letters)*

The group who had this discussion were all professionals in their 30's and 40's. They are the "successful" products of the Spanish academic system with its strict hurdle race of tests and exams. If they feel like this, what do the "rejects", the "failures" feel?

The Exam Taker's imposed Frame of Mind
We have so far had a look at the way EFL testing literature avoids dealing with the exam as a psychological crisis that can generate, stress, anxiety, fear and even panic. We have also looked at clear cases of exam takers entering the testing room in a far from optimal state of body, heart and mind.

But there are other more cognitive aspects to most tests that need looking at. The majority of candidates go into a language exam in a "mistakes avoidance" state of mind. They often have a strong mapping of what they do **not** know or are unsure about and are determined to hide these areas from the examiners. A dramatic example of this came up when UCLES (The Cambridge exam authority) did an analysis by nationality of mistakes being made at First Certificate (FCE) level. They discovered that Japanese students had made no mistakes with relative clauses. They smelled a rat and had a close look at the Japanese scripts - this national group had scrupulously avoided using any relative clauses!

Is a "mistakes avoidance" strategy a resourceful state of mind and heart? Is it conducive to showing your paces, to really shining in the target language? My own feeling is that the fear of falling into error fiercely inhibits natural linguistic and intellectual creativity.

I remember once showing a long letter I had received from a student to the Secretary

of a Language Examination Board. He read through the eight lower intermediate pages of hand-writing, in which the writer was desperately trying to teach me some economics (her speciality) and then looked up and said , pensively, "This text was not written to be corrected." He was dead right. This student wanted me to understand her meaning, despite her language having more holes in it than a piece of crochet work. The exams man was amazed to read a piece of **communicative** writing. In his work he would normally only see **mistakes avoidance** writing.

What do we think we are measuring if we put the exam taker into a linguistically defensive state of mind and then evaluate her shrunken production?

John Fanselow in *Breaking Rules* (Longman, 1987), points out that it is the tester who always initiates, by setting a composition title, by generating a cloze procedure, a C test, a Multiple Choice exercise or whatever. As Fanselow puts it, the test taker is perpetually playing on the away ground, working within a frame strictly prescribed by the other. The candidate is the uneasy guest at the examiner's table.

Sometimes an exam taker refuses to act out the passive, reactive role assigned to him. This was the case in a University physics test where the candidate was asked this question:

Show how it is possible to determine the height of a tall building with the aid of a barometer.

The student suggested lowering the barometer from the top of the building to the street on a rope and then measuring the length of the rope. This logical and feasible answer earned him a zero mark. He appealed. The external examiner, who was brought in, asked him to re-answer the question, giving him six minutes to do so. The student offered this, as one of several possible correct answers:

Take the barometer to the top of the building. Drop it and time its fall with a stopwatch .Then using the formula $S=\frac{1}{2} ar^2$, calculate the height of the building.

The external examiner gave his second answer nearly 100%. The candidate then offered three or four more solutions to the problem, none of them the conventional answer the original examiner had been after. This lad was not in a mood to give

the examiner what he knew he expected. He was determined to play the game on his own highly intelligent home ground. He rejected the intellectual state of obedience and passivity that the exam implicitly required. (The barometer story was written for The New Yorker by Alexander Calandra, professor of Physics at Washington University, St Louis, USA).

Where are the humanistic voices in the world of EFL testing?

If you search the literature for major works on testing by members of the humanistic language teaching movement you don't find much. People like Caleb Gattegno, Earl Stevick, Charles Curran, Lozanov, Herbert Kohl, Gertrude Moskowitz, Bernard Dufeu, John Morgan, Herbert Puchta, Alan Maley, or Alan Duff are fascinated by the processes of learning language. They have written thousands of pages, between them, on the learner as a whole person, as a creative mind, but nothing major springs to mind from their work when we look at the area of testing and exams. (The work of John Fanselow is a serious exception to this generalisation.)

The humanistic movement's failure to address the problem of testing is a grievous one, as no teacher has ever moved through her career without somehow coming to terms with this difficult area. At Pilgrims, with a network of excellent, humanistically motivated teacher trainers, we have offered a course on testing only once in a quarter of a century's work. A cop-out? Yes, I have to admit it is. The area of testing is far too important to be left to the personality types who naturally gravitate towards wanting to measure, to quantify, to evaluate and generally to establish themselves as the gate-keepers.

Some features of an ideal humanistic test

The first question to be asked when testing language is "What **is** language? Following Dufeu, (*Teaching Myself*, Oxford, 1994) I would suggest that language is **Being** rather than **Having**. In my own case I have Latin. I studied it for 8 years and if I have to produce any, **I construct** it, consciously applying the rules I learnt.

It goes something like this: *agricolam (accusative case of "farmer", fourth declension noun, and it can come at the start of the sentence even though it is the object) puella (puella, or girl is the subject, so no "m" at the end) amat (yep, amo, amas, amat, so this is third person singular... and it looks good to have the verb at the end, not like in Church Latin, where it can go in the middle,).* So, the girl loves the sailor.

I hardly need to point out that the way I know Latin has nothing to do with being able to communicate in a language. I know no Turkish, and yet the sounds of *Merhaba* have a place in my head and my heart. *Merhaba* evokes a first meeting with someone, a feeling of beginning and seems to me to be an excellent way of greeting some one. I **am**, **I exist** in and through Merhaba, while *agricola* is an intellectually dead translation of the English term *farmer*.

Merhaba is a Mario word, a Mario pleasure, a Mario handshake. *Puella* is a counter on a language chessboard and does nothing to evoke the many *puellas* I have met and appreciated, in some cases loved. The "signifier", in the case of *puella*, is a thousand miles from the very important "signified".

Following the work of Carter and McCarthy at Nottingham University I would say that language is essentially **relational**, a bridging between two or more people, a central aspect of their coming together, of their meeting. Let me give you a detailed instance of how the grammar of spoken UK English encodes for **relationship**.

If a speaker says, *"She was saying they're coming tonight"* the speaker, by using past continuous, implies that he knows the woman he is reporting.

If the speaker says, *"She said they're coming tonight"*, then we know nothing about his relationship to the woman whose words he is reporting.

This is one of the nitty-gritty examples from the Cancode Corpus of oral English that Carter and McCarthy have been working on for the past ten years.

The trouble with almost all tests is that they deal with language as **having**, as an inert mound of knowledge. Nearly all written tests are **non-relational**, in that the candidate is not doing them in any strong I – thou frame. When the class-teacher sets the test the students is in some sort of relationship with the teacher, but really more with her red pencil, with her language-critical faculty, than with her as a person.

How, then , can we test language as **being** and language as **relationship**? This is

a revolutionary question to which I can only offer a couple of tenuous answers which I have not yet checked out in the reality of an evaluating situation.

Testing intermediate Writing

1. Tell the candidates that the best six pieces of their writing will go up on the school web site and so will be read by other students, by parents and prospective parents. (You are providing the test takers with a real audience, a palpable audience and a largely well-disposed audience, that could well include their own family.)

2. Give the students four or five 1-page extracts of excellent, simple English prose. Ask them to read and re-read these for 15 minutes before writing. Ask them to enjoy and soak up the voices of the writers.

3. Ask the students to write a piece of their own, under the influence of the style of one of the passages. They can even write a continuation of the passage of their choice, or what went before it.

4. The pieces of writing from the test go up round the walls of the classroom for all to read – the students' task is to pick the six pieces to go up on the website.

5. The teacher then does her normal marking according to normal linguistic criteria and awards her technical L2 correctness marks accordingly.

Testing low level writing

The teacher asks each student to write her a two page letter about a topic that has not yet been discussed in class (the topic could be technical, personal or whatever). The student is to write the letter as much as possible in English but is allowed to code switch to mother tongue where absolutely necessary.

After marking the letters, you can usefully ask the students to work with colleagues and try to find adequate English for the mother tongue parts of their letters. This type of test not only permits evaluation but also immediate further learning. The

permission to use L1 allows the students to express themselves in much less curtailed language and so to enrich what they dare to want to try to say.

In both the tests proposed above the candidate has an addressee or audience to write to. Her writing is **relational**, whether addressed to the teacher personally or to the school's website audience. In the first exercise the student is also in a strong linguistic **relationship** to the authors of the model texts. In both tests, the student chooses the topic area to write about, within the relationship s/he perceives with the reader, so in John Fanselow terms, the candidate is playing on her home ground.

To get a good technical mark the student will be aware of mistakes avoidance but also has the human motivation to express herself fully to a reader/s. To say that these two tests do away with exam stress, anxiety and fear is to claim too much. My hope is that they may reduce these negative factors.

Past experiments in humanising testing
The Cooperative Language Movement Tests.
In this approach, widely practised in US secondary education, the students do most of their work together in groups of four to six, and each group is organised to be as heterogeneous as possible in terms of race, class and academic ability. When the time comes for the test, each member of the group of six does his or her preparation together, with the stronger ones helping the weaker ones. It is in their interest to do so, as the students know that, while they will take the test as individuals, and while their test papers will be evaluated individually, the mark they finally receive will be the average mark for the group.

This mode of testing raises the hackles of people in very individualistic societies, for instance Germany, but is realistic in terms of what happens in later life. If a team of engineers build a bridge, the whole group will be judged on the outcome and the less good professionals will benefit from the presence of those who are stronger. Isn't the team you work in judged as a whole, as well as sometimes individually?

Learner-Teacher Co-evaluation
Evaluating another person's work puts you in boss/parent position over them. There have been many attempts at power-sharing over past 50 years and a recent one, at

upper secondary level, is described by Christoph Ruehlemann in his article: *"Sharing the power: action research into learner and teacher co-evaluation"*. (You can read the whole article at www.hltmag.co.uk under Major Article, Year 4, Issue 1, January 2002.) In describing his experiment with co-evaluation in a German state school, Christoph describes a system of careful checks and balances. The first text in the exam is marked by both the teacher and a peer-evaluator, using the same type and number of criteria. They each have a 50% say. The second text in the test is marked for one criterion by the teacher and for three by the peer-evaluator, thus giving the student a 75% say. The third text is marked only by a peer-evaluator, gving the student full power of decision.

Christoph, at the end of his careful, detailed article, asks, *"Do teachers and learners benefit from co-evaluation?"* and then has this to say:

*The answer is a clear yes. The obvious benefit for the teacher lies in the diagnostic exploitability of rating disagreements. ......Astonishingly, **accuracy** turned out to be an area of relative rating harmony between teacher and students...... There was much greater rating disharmony around the criterion **variety**. It became evident that this criterion had not yet been sufficiently well taught and learnt, an insight that contrasted sharply with the teacher's expectations. So, investigating these rating differences may greatly help identify learner weaknesses and define areas of additional learning and teaching.*

Co-evaluation provides an occasion for genuine learner and teacher cooperation in a field where, traditionally, teacher autonomy is paid for by teacher isolation.

Co-evaluation benefits learners too. Getting to read their classmates' texts puts them in the place of the audience, which establishes writing as a communicative act- rather than a language exercise. Interestingly, for learners to accept their peers as 'real readers' it is prerequisite that evaluating and grading is not the prerogative of the teacher, but shared by the classroom community.

Finally, co-evaluation greatly contributes to learner autonomy and responsibility.

Student – self evaluation
In *Freedom to Learn for the 80's*, (Charles E. Merrill, 1983), Carl Rogers describes

the pioneering work of Dr Herbert Levitan, a lecturer in neurophysiology. In the context of an undergraduate course where the contents and manner of teaching were extensively negotiated with the class group, Levitan decided that the marks awarded for the course should be based entirely on student self-evaluation. Each student had to submit the following:

- a portfolio of all written material s/he had produced over the semester
- a diary of reflections on his work over the semester;
- the grade he awarded himself and a justification.

Levitan writes: *I reminded them that I reserved the right, and indeed felt the obligation, to give them feedback on the grade they assigned themselves. I made clear, however, that I would respect their final decision on the grade they wished to have submitted to the University.*

Here are two of Levitan's students' self-evaluations:

Evaluating myself is difficult, but I will try and be objective. I feel I've come a long way since the start of the course. Instead of just learning facts I learned how to ask questions and approach a problem... but more importantly, I learnt how to discover more on my own. I believe my effort in the course is worth a B.

Based on the amount of time I spent in class compared to the amount of time I could have spent and the number of concepts I could have learned I give myself the grade of C for the course. I do not think a higher grade is justified, simply because I did not make a formal attempt at synthesis of a topic of interest (term paper). Also a lower grade than C would not reflect the amount of time I placed in the course and my satisfaction with what I learnt.

Levitan reports that the distribution of self-evaluation grades for the course was:

33% A
45% B
20% C
 2% D

On many previous courses on the same topic, which he had taught without consulting the students on **what** they wanted to learn and **how** they wanted to learn it and without asking them to self-evaluate he had suffered a drop-out rate of 30-40%. On this course no one dropped out.

Yes, of course, Levitan's experiment would not work in all contexts and in all cultures. The generalisable value of any experiment will be constrained by major cultural and belief variables.

Self evaluation of this Article

*   For those readers who are convinced, at belief level, that the psychological aspect of testing **must** be ignored, because otherwise one simply enters an issueless touchy-feely jungle, Mario will have confirmed and hardened their conviction. From now on they will devote yet more energy to their validities and their reliabilities.

*   For those readers who have generally accepted that current testing ways are simply a given about which nothing much can be done, you maybe have half-opened a window on a hazy, new thought-landscape.

*   For people who feel that most of current testing is psychologically unfair, this article may articulate things they have always felt and suspected.

People looking for alternative ways of testing may be motivated to try out the practical systems outlined in the second half of the article.

I give myself a B+++ grade for effort
         a B – grade for breadth and depth of knowledge of the area
         a B + grade for trying to find an appropriate voice for this piece.

(A **thank you** to Ahmet Sofuoglu, Longman Turkey Universities Consultant, for having encouraged me, pushed me, challenged me into giving a plenary on Testing at Marmara College, Istanbul, during their April 2002 Conference, testing being an area that I normally shy away from.)

Mario edits *Humanising Language Teaching*, Pilgrims' webzine for language teachers www.hltmag.co.uk. Mario is a founder member of the Pilgrims trainer network. He has also developed a Cdrom called *Mindgame* with Fletcher de Tellez, Clarity, Hongkong. His most recent book, *Using the Mother Tongue as a Resource in the FL classroom*, written with Sheelagh Deller, came out with Delta in November 2002. In February, 2002, Delta brought out Mario's *Humanising your Coursebook*.