TESTING OR OWLS, PEBBLES AND FANS

by Mario Rinvolucri & Pilgrims

Self-testing seems to be a major part of all thinking activity, and is directly connected to the human ability to set up hypotheses and then check them out. Take what I am doing now, writing sentences for you to read. Parallel to putting my thoughts into words via my fingers on the keyboard, I keep on reviewing what I have just written, what the next sentence is going to contain, the inter-relationship between these pieces in my own mind and my guess at how you might perceive them. Bereiter and Scardamalia (1981) offer the following list of possible writer self-evaluative thoughts:

- the reader may not understand this
- the reader may be offended by this
- the reader does not need to know this
- this sounds all right
- this says exactly what I mean
- I am satisfied with this
- this is inappropriate
- the register is wrong
- this thought doesn't link with the previous sentence
- this doesn't give enough detail
- the confusion is in my head, not in the mise-en-page
- this is too direct/indirect ....

I am certainly experiencing most of these self-evaluations in writing this article.

Evaluative feed-back to self would appear to be an inevitable part of most things we do. Driving a car is another clear example. In the UK advanced driving test the candidate verbalises her self-evaluation as she drives. She explains what she is going to do and has done and why, as well as what she might have done differently. This is simply a making conscious of a normally semi-conscious or unconscious process.

Continous hypothesis formation and hypothesis checking fills the waking life of a person between one and three learning his mother tongue. A person at 22 months says 'ticker' as his version of the word sticker. A few months later he starts sometimes and then firmly and always producing 'sticker'. For this to happen we have to hypothesise that there is a feedback process continously at work, as the child moves from one version of a word to another and then to another.

An approach to second language learning that closely bases itself on hypothesis formation and checking is Silent Way. In a Silent Way beginners' lesson I will pass the little wooden cuisenaire rods round the group so that people can feel them, drop them, listen to them and smell and taste them if they so wish. In reaction sessions after such a beginner's lesson someone in the group will say: "I'm not sure what the word 'rod' really means - maybe it means piece of coloured chalk." All through a typical Silent Way lesson the students are making language-plus-reality hypotheses. Take a Silent Way teacher of Greek introducing numbers. She writes 4 on the board. She mimes a request for the word in Greek. Silence. Student A, after some thought, produces a hesitant TETRA. First hypothesis. The teacher gives feedback silently showing that TE is the correct initial syllable and that RA is the correct final syllable, that the middle T needs replacing by a new syllable. The teacher may well mime this by establishing TE on her ring finger and RA on her index finger. She then mimes a request for the middle syllable by pointing to her middle finger. The students may guess their way to SE or the teacher may decide to loudly and clearly give it. (The teacher works 'backwards' from her point of view so that it is seen as forwards from the students' side.)

founder and popularizer of the Silent Way

What Gattegno is proposing is a socialisation of the natural, internal feedback mechanisms that each language learner possesses and uses. The teacher speeds up the learner's hypothesis formation and checking process by narrowing the speculative range and offering helpful clues. In the example given above the teacher picks the good bits out of the hypothesis TETRA and so organises them that reaching TESERA is a relatively easy task. The teacher in a way 'cheats' by making student speculation reach the correct conclusion much faster and more smoothly than it could unaided.

The whole of Silent Way practice is built around the model exemplified with the 'teaching' of TESERA. Since the teacher is invading a process that is normally hiding within the student and is socialising it, it is vital that the teacher should behave as nonjudgementally as possible, should intrude as little of her own separate personality as possible into the operation. Neutrality is an unattainable ideal. But the reason for its desirability is clear. Hypothesis formation within the mind of the learner is a technical, cognitive process, not a value-laden one. Within Gattegno's socialised model, teacher feed-back should remain as far as possible on the same 'cool' level - neither hot nor cold, so as not to emotionally disturb the thinking learner.

All the above goes to show that testing is a vital and normal part of thinking, be it in areas like writing , driving a car or learning a second language. The testing, that is, that one does of one's own processes.

The problems start when the teacher tests learners to 'objectively' assess what they have learnt, when learners start to compare themselves with those around, rather than with themselves a minute ago, and when the state starts to set examinations that are designed to make some students, often a majority, fail.

State examinations are a major teaching problem: Fanselow in Breaking Rules (1987) suggests that in the 1980's many North Yorkshire schools spend one day per week for the entire school year going over test items. In Nigeria and Somalia, during the last few months before the examination all class time is devoted to test practice. He notes wryly: "When we remember that the idea of a test is to sample what people have been doing, it seems that the emphasis on preparation for tests defeats the entire purpose of sampling."

From a humanistic standpoint what can we say about the political institution of the state exam, the gate-keeping device that ensures everywhere in the world that though many are called few are chosen?

The first thing is to note the gross unfairness of the system. Major exams have a very different psychological effect on different individuals. In my own case exams often filled me with a feeling of adrenalin pumping, joy at

performance, a feeling of challenge and exhilerating risk. The effect they had on my brother was totally destructive - his writing hand trembled so much at his exams that he could hardly hold a pen. He passed only one subject and this 'failure' governed the path he has taken through life. I have no reason for doubting that his intelligence was greater than mine, this being the judgement of a person who taught us both. The British State's academic decision about Bernard and me at 16 was grossly inaccurate because it put him in a situation he couldn't bear and offered me an ideal ring to show off in. I jumped through the hoops with more glee than awareness or dignity.

Scanning books written on EFL testing one finds little written about the psychological reality of the examination room. Most specialists in testing seem to be much more worried about making sure they test the right things in the cognitively right way than wondering about the personal reality of the examinees. And yet this reality determines whether the exam is an accurate measure or not. It takes a plunge into journals of psychotherapy to find educators willing to think realistically about candidates' feelings about exams. Workers like Beck have produced desensitization videos to help people with exam phobia. The candidate views a series of anxiety provoking scenes, graded according to the intensity the scene is likely to arouse. When each scene is combined with a relaxation exercise the person no longer feels the same degree of anxiety as before. Beck offers nine scenes to exam phobics. Here is a selection of them:

2. A person tossing and turning the night before the examination is to be taken.

5. A typical classroom with students talking nervously before class. The instructor enters, carrying the examinations.

8. Close-up of time slipping by as the anxious student writes frantically on official paper.

(Beck TK Videotaped scenes for desensitization of test anxiety - Journal of Behaviour Therapy and experimental psychiatry 1972, 3, 195-197).

My brother could certainly have done with some of Beck's work, but one is led to wonder why the State should set up a system so unfair as to drive the people subjected to it to needing therapy. Desensitization techniques are mostly used with people who have phobias, like fear of flying, claustrophobia, terror of spiders, etc.

To see that exams are grossly unfair in not taking into account the situation of the candidate on the day, one only has to think of 500,000 sixteen year-olds taking a State exam. 250,000 of these will be girls. 50,000 of them will be having their period on the day of the exam. A sizeable percentage of the 50,000 will sit the exam in considerable physical discomfort. And people go on about 'objective testing'. It's absurd.

One of my proficiency students last term told the class, when we were discussing the exam, that all you had to do was provide the examiner with what s/he wanted/expected. She explained that she had learned this early on in her career at a German gymnasium. One of the oddest things about exams, as Fanselow points out, is that the examiner nearly always initiates while the candidate is perpetually playing on the away ground. The point is amusingly illustrated by Alexander Calandra, professor of physics at Washington University, St. Louis, in a New Yorker article:

Calandra is called in as a referee by a colleague who wants to give a student zero for his answer to this physics question: "Show how it is possible to determine the height of a tall building with the aid of a barometer." The student suggests lowering the barometer from the top of the building to the street on a rope and then measuring the length of the rope.

Calandra has the candidate re-answer the question; he gives him six minutes. The student suggests this as another of many correct answers:

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

The student was granted nearly full credit for this second answer. He offered

three or four more, none of them the conventional answer the examiner was after. This guy was not in the mood to give the examiner what he wanted. He was determined to play the game on his own highly intelligent home ground.

With external State exams all you can do perhaps is palliate the harm they do your students. What about tests you do have control over? Here it is worth looking at some of the solutions radical thinkers have proposed and the rest of this article focuses on ways of making evaluation less unfair and fraught.

One sensible objection to most testing is that it isolates the individual from her peers. In most situations where you have a problem to solve you turn to a peer for help. The doctor asks for a second opinion, if I can't move a piano I get someone to help. In some EFL classrooms today students are encouraged to help each other on a day-to-day basis, but woe betide them if they continue this co-operative behaviour on the day of the test. It's called cheating. How then can the individualised test be socialised? ~~One suggestion comes from Japan (reported by Vincent Broderick).~~ In preparation for the test students are organised to revise in threes. In each three there is a strong student, an average one and a weak one. Each person takes the test individually, without help from his/her partners. Each person receives the 'average' mark for his/her threesome, so if the strong one scored 90%, and average one 60%, and the weak one 40%, each person in that team would get 63.33%. The clear intention of this form of testing is to make sure that it is in the interests of the strong to really help the weak during revision. It is in their egoistical interest. In some parts of Northern Europe teacher reaction to this proposal has been very fierce, with people angry about bright children being penalised and the offence to individualism. Not a Thatcherite vision of testing.

A solution to testing suggested by my Pilgrims colleague, Jean-Paul Creton, is to have children make up their own tests. He divides his class into five sub-groups. Each is responsible for preparing a fifth of the total test. They prepare the questions and decide on the scoring system. Jean Paul helps each group correct their section of the test. The next day the whole class sits the test. The setters then score their sections, and the work is done. Jean-Paul is rid of a major chore and the kids have really revised their work by the act of preparing the test. Even the weakest people do reasonably well

One suggestion comes from the cooperative learning movement in California as reported by David W. Johnson in Circles of Learning (1984)

on the section of the test their group set! A solution to testing which gives power to the children and at the same time reduces the teacher's work load is a methodological jewel.

cooperative learning movement

The ~~Japanese~~ testing system and Jean-Paul's are both rather radical solutions to the problem of testing. A 'gentler' one comes in Veronica Smith and Christine Klein Bradley's book on teaching translation: **In other Words - Arbeitsbuch Ubersetzung** (Hueber, 1987). The authors suggest that university-students, before sitting a translation exam, should present their teacher with a passage in one language and three different translations into the other language, ranging from fairly literal to much freer. They ask the teacher to mark each translation. The authors provide the students with this "teacher-testing" passage and the three translations! The idea is to invite the teacher to lay her cards on the table and to show students how she expects a translation to be done. Students may sometimes already have a pretty clear idea of their teacher's fads and prejudices but the point of the Smith and Klein Bradley proposal is to force the examiner to make her criteria publicly known in an area like translation, judgements of which are inevitably highly subjective. It's a way of making the judge/teacher read out the law under which she is either condemning or acquitting the candidate/criminal.

(1987)

John Fanselow suggests substituting observation for testing:

"During the five to twenty minutes needed to develop one multiple choice item - without pre-testing it - five to twenty student communications can be written on note cards. If the communications are made during a break by students to each other (EFL students in an English speaking environment), or are made as the students are engaged in the actual solution of a real problem, such as the putting together of a jigsaw puzzle, or in a conversation of their own, the communications are likely to be more similar to those they ordinarily make than responses to questions in tests. They may be less inhibited as well, and so a more valid picture of capability may result".

Fanselow is rightly worried about the idea of prejudicing people's language performance simply by putting them in the 'test' frame of mind and body. It's a bit like asking crack marksmen to prepare for competition by having a swim

in heavily chlorinated water. He writes:

"In observing oral interview tests, I have noticed that nine times out of ten the people being tested keep their hands at their sides or held together in their laps. The body is usually held very erect, much as it is during a classroom recitation".

The most radical proposal I know for coping with the problem of grading students is to ask them to do it themselves. In Freedom to Learn for the '80's (1984) Rogers describes the pioneering work of Dr Herbert Levitan, a lecturer in neurophysiology. His course for graduate college students had a reputation of being a good one but very demanding. The thing that worried him was a student drop-out rate over the semester of 30-40%. After six years of traditional teaching during which he took all the responsibility for syllabus, method and testing he decided to negotiate content with his students and ask them how they wanted to work. In the area of laboratory experimentation, for example, he asked them whether they felt they would learn more from experiments which:

1. were failsafe - guaranteed successful
2. had a 75% probability of success
3. had a moderate probability of success
4. had a low probability of success
5. were impossible, that is to say where no success was probable the first, second or third time, but after a dozen experiments the success rate would reach 50-60%

Most students picked the third category though many felt that the fifth category was the closest to the conditions in real research. Category 1 is the usual lecturer choice.

A major student worry through the course was how the evaluation of their work would be done: "the students repeatedly asked how I was going to evaluate them, a question that reflected their past experience in courses in which their ultimate objective was to please the instructor".

Levitan decided that the marks he 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 he had produced during the semester
- a diary of reflections on his work over the semester
- the grade he awarded himself and a justification

"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 student 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 learned 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 would not reflect the amount of time I placed in the course and my satisfaction with what I have learned".

Levitan reports that the distribution of self-evaluation grades for the course was: 33% A, 45% B, 20% C and 2% D.

The major success of the course was that the drop-out rate was reduced from 30-40% to zero.

To have got students, long accustomed to being marked from on high, to take self-evaluation seriously is no mean achievement. In the early 70's a group of EFL teachers in the University of Valdivia in Chile read Rogers and

decided to try self-evaluation with their students. We used a mixed system:

- test
- self-evaluation
- group evaluation of each student
- teacher evaluation of each student

I seem to remember self-evaluation was worth about 10% of the final mark.

With us the student evaluations did not work honestly. Our students were too used to having marks used against them in a ghastly cat-and-mouse game. I suppose many of them thought these 'gringos' had simply gone crazy: who ever heard of students awarding themselves their grades?

Our failure was also to do with introducing self- and group evaluation as fringe items. We did not trust what we were introducing sufficiently ourselves. Why should the students? While Levitan revolutionised his whole power relationship with his students we offered ours a tiny area of power at the end of the course. They found it derisory and they made it derisory.

The reason for reporting this failure is two-fold:

(i) normal honesty
(ii) to make the point that a Rogerian transformation of the evaluation system in education presupposes a major change in the educator, in the students and in their power relationship to her.

I certainly was no way near ready for the kind of revolution in personal relations between students and teacher that Rogers proposes back in 1972-3. Maybe I am beginning to be ready for it in 1989. Slow learning.

The last word goes to the voices from the alternative, peasant hilltop school of Barbiana, Letter to a teacher, 1961.

"Owls, pebbles and fans - no, not for their good. You gave an A in French to a boy who, in France, would not know how to ask where the toilet was. He could only have asked for owls, pebbles and fans, either in the singular or

the plural. All in all, he knew perhaps two hundred words picked carefully for being exceptions, not for being commonly used."

The result was that he hated French the way some people hate maths.

Examinations should be abolished. If you do give them at least be fair. Difficulties should be chosen in proportion to their appearance in life. If you choose them too frequently, it means you have a trap-complex. As if you were at war with the kids.

What makes you do it?

Beck, T.K. Videotaped scenes for desensitization of test anxiety. Journal of Behaviour Therapy and Experimental Psychiatry, 1972, 3, 195-197.

Bereiter Carl Scardamalia M. From conversation to composition : the role of instruction in a development process. In Glaser.R. (ed) Advances in Instructional Psychology. Vol 2. Lawrence Erlbaum Associates pp. 1-64. 1981.

Fanselow John F. Breaking Rules . Longman, 1987

Johnson David W. et al. Circles of Learning , Association for Supervision and Curriculum Development, 1984

Rogers, Carl. Freedom to Learn for the 80's, Charles E. Merrill, 1983

The school of Barbiana, Letter to a Teacher, Penguin, 1961.

Smith Veronica and Klein Bradley C. In other Words - Arbeitsbuch Übersetzung. Hueber, 1987

yes. (these are the authors.