ASSIGNMENT 9
LANGUAGE ASSESSMENT
Presented by: Andrés Melo Cousineau
November 11th, 2003

(All quotations ---except when specified otherwise----- are taken from; Brown, Douglas 2001, *Teaching by Principles: An Interactive Approach to Language Pedagogy*, Second Edition, White Plains, NY: Longman; abbreviated as IALP)

CHAPTER 21 LANGUAGE ASSESSMENT: BASIC CONCEPTIONS IN TEST DEVELOPMENT

In this last assignment, once again, I must apologize to the reader. As most of my philosophical and political education deals with issues regarding ethics, I cannot but decide to avoid answering one of the selected questions found in our distance program. Instead, I will center on one that is of much more relevance to me, and I hope, to the reader and the ESL area in general as well. This is the reason for my omitting the important Question No. 4 ---which deals with oral proficiency and its testing--- and substitute it by Question No. 9 as it appears in Brown's book. A reminder, in the previous assignment a dilemma was presented, the consideration of Hitler as a good writer. That he might have met the formal criteria for writing, and yet be an abhorrent human being, is quite paradoxical. In this respect a concern for ethical question is of the utmost importance for ESL. Perhaps, reflective philosophical thought might provide some light on these difficult and complex issues. Brown's question reads as follows: "… Among familiar standardized tests, what ethical issues might emerge? Does the testing industry promote a widening of gaps between educated and uneducated, rich and poor, 'haves' and 'have nots'?" (pg. 400).

As I believe that ethical concerns are of a more fundamental nature than others, I will begin by considering this question which, in order to avoid confusing the reader, I will number as Question '0'. And numbering it so is quite relevant as regards testing. When taking an exam, have you ever gotten a "zero"? Probably not. But it seems to me that a zero is a sort of nothingness, a kind of emptiness. Getting zeros amounts to not being much. Or so would traditional kinds of exams have us teachers, and more importantly learners, believe.

QUESTION 0

According to Brown, testing –specially the traditional kind---- involves a method whose purpose is that of measuring a person's ability in a given domain. (Brown, 384-5) What one should emphasize here is the indefinite article, it is <u>a</u> method. In this sense it is not the only one or, in some cases, an indispensable one. Furthermore, one need underline that the need to measure people's ability seems to be a completely modern project stemming from the development of the experimental method which lay at the basis of the surge of the scientific revolution in the 17th century. Testing to measure becomes a priority for it is only by testing that knowledge is actually made numerically and statistically analyzable. This is quite problematic, for one could see testing not simply for the sake of measuring. In this sense, it is common for you to say to yourself: "I will test myself". And this does not necessarily mean measuring yourself by means of a quantifiable yardstick, or grade-stick. When you tested your carpentry abilities building the porch, you did not end by saying, "well, I give myself a B+". You might have, if you are a teacher; but in general non-teachers would say something like "it is nice, but too bland", or "I have to improve the corners", or "man, this is tough, as we Canadians usually are"; or something of the sort. Everyday abilities seem to do just fine without ones measuring them.

Or to put in other words, a historical examination of how testing came to be, remains crucial to our societies where standardized testing has absolute importance. Measuring ourselves and our knowledge has become a way of life. By understanding this, we could move towards understanding why our scientifically oriented societies place so much importance on them as "measuring" devices. Foucault is a famous philosopher who has undergone such historical understanding of multiple facets of our self-understanding. For instance, he argues that the very notion of statistics is generated because of the 18th century concern of the state to control its population; hence the semantic proximity, "statistics", "state". (See his lucid *History of Sexuality* Volume 1). The test measures its citizens, and some might regard this as a form of control which generates a certain class of citizenry. If this is true, standardized testing might lead to a certain kind of normalcy, rather than to an emphasis on excellence. Brown knows this and therefore he quotes Shohamy who argues: "Tests represent a social technology deeply embedded in education, government, and business, as such they provide the mechanism for enforcing power and control" (pg. 397). Shohamy seems to be quite Foucaultian. Or let me put it another way. You may not be the brightest kid around –if brightness could ever be correlated with exam

taking--- but your hugs may be the finest. Not to see that much of what is important to humans goes beyond measuring is to suffer from a crippling blindness.

In this respect, the fact that this distance program does not seek to take up this issue as a central one is, at least to me, a little troubling. For you see, as I am writing this assignment, I <u>know</u> that you are testing me, you are measuring me. And I am not quite sure if I want you to measure me, or to test me. Do you not feel ---- maybe like all of us teachers---- a certain power flowing from your pen? But the fact is that the ethical nature of testing is not one of the questions we should focus on in our distance program. I take it it is not emphasized because the distance program is geared towards new teachers who, because they are beginning their teaching careers, should first develop testing abilities in order to ----later on--- have certain criteria to puzzle about them. But this in itself is quite puzzling. For we, I am speaking here of teachers in general, are told that we should test and wait until we are good testers to ---later on--- question that which we have become great at. If habit formations are central to what makes us human, then surely ingrained habits will be tougher to dispel. Similarly doctors focus on testing diagnosis though their careers and, only later on, may actually start to puzzle as to how to develop creative and healthy doctor-patient relations.

However, Brown in truly Aristotelian fashion (see also his plea on page 444), tries to move away from such radicalism and seeks instead to portray not only the benefits, but as well, the dangers of the desire to measure both ourselves and others. This is why I myself will try to put the argument in other, less drastic, terms. The issue of evaluation could be understood by making a parallel with an essay by Professor Carens entitled "Realistic and Idealistic Approaches to the Ethics of Migration". Briefly, from the idealistic camp an open borders policy is morally justified and required. However, from the realistic camp considerations of security and economic change have to be considered as well. In the end, policy makers must balance the two as best they can. Similarly, one could argue that both realistic and idealist concerns must be taken into account in the consideration of the evaluation of learners and colleagues.

It seems as though both approaches seek to answer different questions, and both could, at least to some extent, learn from each other's presuppositions and goals. For example; you might be an adamant critic of exams -----–or better, might have become one through the years--–— and yet you work in an institution which requires testing. The dilemma can only be resolved by seeing the relevance of both sets of approaches. Or you could work in the Ministry of Education

where all too realistic policies shape the future of a country. Let us then look at both approaches, even if all too sketchily.

The main types of questions that issue from an idealistic approach to test-taking would include: Should learning be motivated through exams which are, even if of the alternative kind, mostly extrinsic? If education is guided by examinations, then, in the end how could one know if the learner has been motivated for the sake of learning itself, or for the sake of learning of showing the examiners how much she has learned? Why is it that most students cannot remember much of what they are tested on at the end of a semester? What does it mean to become accustomed to a system which works via examinations? Have you not seen the recent ads by the Hockey League on children criticizing their parents? Should we not be pushing ourselves in radically different directions? Besides, are all our students necessarily educated by being measured? Do we not have some students ---maybe just a few--- who can do without them altogether? Do you think great creative geniuses became creative through test-taking? In the linguistic area, Chomsky surely did not. In philosophy examples abound of a rejection of this correlation of understanding with examination (Nietzsche, Wittgenstein, Plato, Aristotle). Do we not find excellent test-takers who even go on to complete PhDs? Could they not have gotten so accustomed to testing their knowledge, that they might forget to "test" who they are via non-measuring standards? Or to put in other words; how to understand an identity which is created to a great extent through external considerations such as exams? As we argued previously, everyday life seem to do just fine without testing as a measuring device. And more importantly, under the idealistic approach one should consider why tests are so damaging, particularly of course, to those who do not pass them? To fail is to become a failure; semantics becomes all too visible. Moreover, standardized tests do precisely that, they standardize; and those above and below the mean seem not to receive the particular consideration which they should. In this respect, can standardization lead to excellence and creative resolution of pressing modern dilemmas such as those underlying "sustainable development"? Does it not, instead, lead towards greater and greater homogenization, regulation and normalization?

Through the years, I have become more conscious of the problematic nature of testing. That was not always so; the exam matrix held me bound! But not only has my testing ability changed, more importantly, it has done so in some areas more so than in others. The relevance of my knowing whether another knows what I know, has become less and less important. Knowing that another puzzles about certain taken-for-granted issues, that is much more appealing.

Learning that some have come up with creative "solutions" to paradoxical elements of human nature –such as the need for measurable examinations---- is far more enticing still. To put it dramatically; to have you as learner merely repeat what I have said to you, well that I can do at home on my own.

Please do remember I am speaking here form the idealistic approach to testing. Under this perspective, which I have come to defend more and more, the former thorny and troubling questions would be answered all too easily. Well, if the consequences of testing are such and such, then certainly we must move away from such measuring devices towards a more liberal and humanitarian approach to education. And that we should try it out, is certainly the battle cry of the idealistic camp. I have taught courses where at the beginning of the course I would say, all of you will pass, so study because you want to. Most people seemed to enjoy the class more, and quite ironically, for most students the class become even more demanding. Facing your own ignorance is humbling, facing a zero, well that might just lead to a radical hatred of understanding itself (See also the story *The Elephant* by Slawomir Mrozek). This is one of the reasons why we have to take heed of the words found in one of the texts from the already mentioned course on literature I taught to EFL students. The passage is found in an excerpt from the book *Kaffir Boy*. These words remind one of the difficulty we have as children to be led towards the path of education. For most of us learning to love to learn is a struggle. Just recall your first days at school. Here is the conversation of a impoverished mother trying to convince her child on how education changes one and the world:

> "Education will open doors where none seem to exist. It will make people talk to you, listen to you and help you; people who otherwise wouldn't other. It will make you soar, like a bird lifting up into the endless blue sky, and leave poverty, hunger and suffering behind. It'll teach you to learn what's good and shun what's bad and evil. Above all, it'll make you somebody in this world. It'll make you grow up to be a good and proud person. ……….. I always wanted to go to school, but couldn't because my father, under the sway of tribal traditions, thought it unnecessary to educate females. That's why I so much want you to go… Promise me, therefore that no matter what, you'll go back to school. And I in turn promise that I'll do everything in my power to keep you there."
>
> "With tears streaming down my cheeks and falling upon my mother's bosom, I promised her that I would go to school "forever". That night, at seven and a half years of my life the battle lines in the family were drawn. My mother on the one side, illiterate but determined to have me drink, for better or for worse, from the well of knowledge. On the other side my father, he too illiterate, yet determined to have me drink from the well of ignorance. Scarcely aware of the magnitude of the decision I was making, or, rather, the decision which was emotionally thrusted upon me, I chose to fight on my mother's side, and thus my destiny was forever altered" (*World Writers Today*, pg. 331)

Quite idealistic you might argue; I have no problem with that. But more importantly yet, I bet you the *Kaffir boy* has even less of a problem. For these words most of us cannot even understand. We have had education as a possibility, though not an easy one of course. But what is certain is that these words empower education beyond examination. Education can be a powerful transformer. Primarily of ourselves, and by way of our struggles, perhaps of others as well.

But let us move back to reality, some would say. Let us do so. Let us center on a more realistic approach to traditional test taking. Aristotle's *Politics* is famous for its strange ordering. He leaves the last two books, out of 8, to the most perfect and ideal political regime. But before heading in that direction, he bids us consider what can realistically be done to improve less perfect and actually quite unhealthy political regimes. Much can be done for Canada's federation, much for social discord in war-torn Colombia. Aristotle writes: "It is not enough to ascertain which variety of constitution is best for each state. We have also to ascertain the proper way of constructing these –and other—varieties" (*Politics*, Book VI; Methods of constructing Democracies and Oligarchies with a special view to their greater stability" (1317a *6). Maybe considerations regarding testing can learn from this approach as well. In this respect it is not enough to ascertain what the ideal non-test like reality would look like; what "ought" to be, generated mostly from a constant negation of what "is", rather than through concrete formulations. In contrast, one should address as well how to construct actual testing so that it may become more stable and less damaging to teachers and students alike. Seeking certain communicative bridges between both approaches provides us with a more holistic and enriching perspective. Building bilingual bridges becomes a truly important affair.

Ours is no longer a traditional community constituted by thousands of citizens. Our modern states, for better or for worse, are constituted by millions of citizens seeking the benefit of the right to universal education. Education is a right of <u>all</u> and not of a few; even though few might actually want to seek in education their life. Mass education has become a modern dilemma. Ever more crowded rooms constitute the educational space for us. We moderns do not have the benefit of the Greeks with their much smaller political communities in which greater conversational interaction was made possible. It is no wonder the Greeks did not have exams! What other exam except being yourself. Personalized dialogical encounters are reserved , if at

all, to the higher stages of post-graduate education. Fighting this trend represents a daunting challenge to us.

Besides, meeting a minimum of standard criteria becomes of absolute importance if education is the gateway to development and economic vitality. No wonder nations have tests such as the SAT in the USA or it equivalent the ICFES in Colombia at the high school level; or ----for entrance to post graduate studies--- tests such as the GRE and GMAT in the Canada, or the newly implemented ICFES exit-exam in Colombia. Policy makers and parents have great concern for levels of achievement. If Ontarians do poorly in science and math (usually the two most sought after scores), then the alarm bell rings.

Realistically speaking, all too often we hear in our modern times: "well exams are a necessary evil." Something like the famous political expression that most everybody is against war, but it seems as though some wars ought to be fought. It is not for nothing that we celebrate Remembrance Day today. In a similar vein, realistically speaking you may argue against the TOEFL until you are blue in the face, however, if you desire to enter UfT or York University, you'd better be thinking about how to pay and how to prepare for it.

Since the rest of this assignment takes for granted a realistic approach to exam taking, I will not develop its concerns and possible changes within this question. Let me just point out that Brown is too much of a realist to properly present a stronger, more appealing, case for the idealistic approach. Ironically, his defense of bilingualism turns quite uni-lingual here. If measurement is of importance to us, then let us measure. Out of the 40 pages dedicated to testing, Brown dedicates 2 or 3 specific pages to the ethical issues arising from a radical questioning of the nature of measurable testing within our societies. A more balanced diet would serve us well. Let me end this question, then, by recalling Professor Carens' own words on the complex interrelation between idealistic and realistic approaches to the ethics of migration. Justice seems to demand that both be considered under varying circumstances and varying individuals; teachers and professors alike must make informed decisions as to with whom, how, where and when to tilt the balance towards one or the other:

> "These forms of reflection, situated somewhere between the extremes of the ideal-typical idealist and realistic approaches that I have sketched, can make important contributions to our understanding of the ethics of migration. But they do not preclude the need for analyses that are either more realistic of more idealistic. There is no simple correct starting point for reflection, no single correct set of presupposition about what is possible" (pg. 11)

Aristotelian flexibility and the needed for practical wisdom ('*phronesis*') permeate this passage. Nonetheless, in the case of our examination-prone culture, creative defenders of the idealistic approach are more and more required. Maybe then, the self-sustaining and self-empowered realistic perspective, may indeed see the relevance of transforming its all too rigid presuppositions. Perhaps, through this dynamic interplay of forces we may move towards acknowledging that the firmest and most educational of all assessment is that of authentic self-assessment. Learning to assess yourself, rather than having others assess you, can teach you to enjoy learning just because you yourself are the greatest benefactor.

QUESTION 1

If one is agreed that realistic considerations are of some importance in the understanding of testing issues, then one must seek to understand the complex nature of testing itself. The creation of a test is no easy affair, and more importantly the creation of a bad one quite dangerous. So then, what are the primary concerns, once one accepts to move to a more realistic approach to testing? Well one must consider issues relating to practicality, reliability, content validity, face validity and construct validity. As we shall see, these terms generate not a peaceful coexistence between both approaches, but rather many times generate a struggle between them. For instance, there exists a tense relation between "content validity" and "construct validity", as we shall see in answering question number 3. Moreover, the meta-testing terminology itself is utterly baffling to a newcomer. We ESL and EFL teachers might be accustomed to hearing these words, but when reading them one ought to be thinking whether they have become all too far removed from what learning is all about. Consider the technicalities: formative assessment, summative assessment, discrete point test integrative test, criterion referenced measurement, norm reference measurement, diagnostic testing, achievement testing, proficiency testing, progress testing, placement testing, objective testing , direct testing, indirect testing, washback effect, method effect, performance testing, random errors of measurement, systematic errors of measurement, passage dependency, alternative testing, self-assessment, peer-assessment, authenticity, cooperative testing, and the list goes on. The terminology is of such complexity that I bid you read this excerpt from the Canadian Language Benchmarks on what Language Assessment is. Please try to read it, imagining your are not the ESL expert that you are:

"Language assessment" is often used instead of "language testing" both for the purposes of placement (placing learners in appropriate programs) or achievement (assessing learner outcomes against program objectives). At the end of each stage there is a suggested performance criterion to be used as a point of reference to monitor the progress to the learner in a specific Benchmark. This is neither to be confused with nor used as an assessment test or evaluation test" (pg XIII)

But, it is written for experts, you might tell me. Yet, that is the point; are experts sometimes not the most difficult to convince about opening their paradigms for they have so specialized that, even to enter their conversations, one must have undergone a linguistic transformation? (Please recall what I have argued in all previous assignment on the self-enclosing nature of specialized paradigms). Sometimes it seems as though you need a PhD to "test" somebody.

As you can see, the idealistic approach seems to creep up on me without my even noticing it! Pushing it aside, let us then focus on the really important one, the realistic approach. Under its paradigm the considerations to be taken into account in the making of a good exam include:

1. practicality: Exams ought to be practical; just in the wording one is to understand that idealistic approaches must give way to practical considerations. And this is absolutely correct; though absolutely troubling if undertaken without a constant reference to idealistic approaches. Mass education requires mass means for testing. Recall your first years as a university student. Being 18 and arriving to Canada as an immigrant who was not an immigrant for I had a Canadian Passport but had never lived here, I began my studies in Political Science. My Intro to Political Science was certainly one of the most shocking experiences. As I entered the packed conference room, I could only see rows and rows of young faces awaiting a minute teacher --- way down there---- who even had to use a microphone to make himself heard. I surely remember nothing about what I was tested on; but I do remember the practical considerations regarding exams at the end of the semester. The first exams at McGill University were presented in the multiple gymnasiums of the university. The practical aspects of setting up this complex testing reality are baffling. Production of exams, exam supervisors, exam materials, costs of exam correction, payments to T.A.s; all in all, an awe-inspiring practical feat.

Or take the case of the CCA and its quite impressive capacity in comparison to many English teaching institution here in Canada or abroad. Its 5000 monthly students presented exams the very same day for all 12 levels and 4 advanced courses (the system has changed a bit since I left). Practically speaking, setting this up must be an administrative nightmare. How many

people were involved? Secretaries in charge of counting and numbering the exams for each classroom; once the 25 copies for each classroom were manually numbered, they recounted them so that no security issues would arise. Setting the correct tapes with the corresponding exam package. Handing out of exams at a specific time in over 50 rooms, both downtown and up north. Teachers trained to carefully check that students not take the copies to their houses. Learning to use the grading formats provided for what were quite excellent exams (though the history of examination at the CCA is <u>highly</u> complex). After exams had been corrected, they had to be quickly returned and counted appropriately once again for security reasons. Were you to lose an exam, it would cost you where it hurts the most; in your pocket. Each lost exam cost around 100,000 pesos (a lot of money). Furthermore, you had to come up with another exam format which could take several weeks to create given their admirable complexity. Then, handing in of exams to the personnel who had begun the whole process two days before exam day. And finally, a few days later, peer-correctors checking on different teachers to see if scoring was dependable or not. All this in a period of just three days. This is quite amazing. Idealistic approaches have much to learn from this.

 But from the idealistic perspective one could retort; in the long term will not such practicality do away with crucial elements of what a more creative solution-oriented type of learning requires? Terminologically speaking would not what "practical" is taken to be here, need to be placed on the table and set up for discussion? If by "practical" you mean short-term consequences, then for sure follow realism all the way. However, if by "practicality" you mean how education will be able to develop practical solutions to the complex practical problems of the modern world, then truly a more balanced diet is required. One such critique writes:

> "These tests are convenient and relatively inexpensive ways to grade students and schools. The danger, critics contend, is that we are putting testing ahead of teaching and relying too heavily on one form of measurement at the expense of others. In our zeal for assessment, we may be distorting what children learn and not preparing them fully for the challenges of this new century. If classroom time is devoted to drilling students on how to answer multiple-choice questions and write three-paragraph essays, the ability to memorize facts and quickly fill in answers is rewarded.
>
> Meanwhile, skills essential to navigating our increasingly complex world — thinking creatively, building teams, challenging assumptions, conducting research, sifting through conflicting data — are at risk of getting short shrift because they are harder to answer in the formats of standardized exams.

"They [the tests] promote a rigid, formulaic, rote thinking style," argues Peter Sacks, the author of *Standardized Minds: The High Price of America's Testing Culture and What We Can Do to Change It,* published this year by Perseus Books. "The testing culture promotes thinking that is contrary to the innovative, critical, creative thinking we need to solve problems." **28 November 2000 Illinois Issues** www.uis.edu/~ilissues

Problem-solvers, or better, puzzle-confronters, versus test-takers; that may perhaps be the most practical problem we have to face in the devising of sound and healthy educational programs.

2. reliability: from a realistic approach, reliability is absolutely central to test formation. Tests must be both consistent and dependable. On the one hand, consistency is gained when the test is given to the same person, or matched subjects, on different occasions. If the results are similar, then one achieves consistency. If you have had to test many learners, one knows that a good exam usually gives out the similar range of good, regular and bad results. Few in the borders areas, many in the middle "standard" areas. Consistent exams will provide very similar results. In this respect, it is usually the case for a beginner teacher to be utterly shocked to see that all her efforts have resulted in a considerable part of her learners "failing". In shock she might ask herself (I certainly did) "But, how could that happen if I am such a good teacher?". Nevertheless, through time what is quite shocking indeed is that one grows comfortably accustomed to the consistency of results. You already know that most will be in the median ranges and a few over and below. So, why do much about it?

  From the idealistic perceptive seeking consistency is not without its problems. Just as the law is blind to particularities, so exams which are quite consistent can be consistently unhealthy for some otherwise quite brilliant individuals. I have known of students who truly were just terrible exam-takers. Coached and all, they would still, sometimes by a few points, fail. Again, that the TOEFL is consistent does not preclude the troubling consideration that a native Aboriginal should have to take it to enter University. Why do you not learn Cree instead? Or in defense of all this terrible test-takers, whose life depends on the consistency admired by others, one can make measurements acquire a real, touchable, and unique face. This is the face of Jason:

> **"J**ason Braham earned steady A's and B's as a pupil at James Madison Elementary School in Chicago. He won an award for nearly perfect attendance. But when Jason took the Iowa Tests of Basic Skills last April, he choked under the pressure and his scores on the standardized exam slipped from the previous year. As a result, Jason had to attend summer school and wasn't allowed to participate in his eighth-grade graduation."

Jason is a person, the son of a mother. For the idealistic approach no examination can consider the particularity which makes of you you, and not simply another me. Let me put it like this. If we are to seriously take the issue of education as learner-centered, then a defense of consistency is quite inconsistent under many circumstances. Justice cannot see particularity, the changing, the unpredictable nature of much human activity. But ethics, from the Aristotelian perspective, does in fact demand that such particularities be taken into account. As he writes on a passage dealing with the all or nothing nature of the law:

> "but it is not difficult to understand the instruction that the law gives … but how actions are to be performed and distributions made in order to be just ---to know that is a harder task than to know what one's health requires; because in medicine too, although it is easy to know what honey and wine and hellebore and cautery and surgery are, to know how, and to whom and when they should be applied to produce health is no less a task than to be a qualified doctor.(*Ethics* 1137a10)

For the sake of our learner's health, greater sensitivity is required in order to be able to actually see dynamically changing circumstances which require that exams be transformed for some students under certain conditions. Alternative healing and alternative testing seem to share much in this respect.

   Under the heading of reliability, the second concern one comes across, is that of dependability. A reliable exam must have consistency not only as regards its takers, but likewise as regards peer graders. One sometimes assumes one's way of grading is universal. This is far from the case. In Colombia students will say of a harsh grader: he is a "cuchilla" ("a knife"). This cultural information should be quite shocking to you. At the CCA, as part of our training, we were provided with multiple exams which we corrected and then compared with other colleagues. But more than this, it is actual comments by students which reveal how diverse teachers seem to correct. Lack of dependability within a curriculum seems to upset learners. They cannot understand how 'x' passed with teacher 'y', while they themselves worked much harder and yet failed to pass under teacher 'w'. Or take the following response to a question in an exam where each response is worth 3 points. To the question "What did she teach", a student may write down:

Andrés Melo Cousineau

"She taught man to dance".

Different teachers will reduce differently. For some it is utterly unacceptable to misspell "taught", and therefore they will take off 3 points. Others may feel that the student has gotten the message through, so they will only take 1 point off. Still others will reduce following other subjective considerations.

But then one need ask, how can an institution generate greater dependability if such diversity cannot be fully homogenized? In this respect greater experience seems to provide important elements for distinguishing between dependable and non-dependable exams. Besides, most institutions will in fact provide you not only with an answer-key which includes the general guidelines for correction (e.g., 2 points off for verb error , or something of the sort) but also peer-correction after exams have been already handed in. These will help bridge the subjective gap.

3. <u>Validity</u>. Herein lies by far the most complex of all the criteria surrounding exams, it deals with the "degree to which the test actually measures what it is intended to measure" (pg. 387) A written test is as invalid for a competing athlete, as a multiple choice exam is for philosophy students. Validity requires aiming at testing the skills required for a given domain. Little philosophy is developed through multiple choice exams because reflection is not simply a matter of selecting from pre-given alternatives. Rather, philosophical validity arises when learners become all the more autonomous in their reflective lives. In other words, validity regards adequate and relevant measuring, as opposed to just <u>any</u> kind of measurement whatsoever. But measuring knowledge, that –if one is open minded--- is a tall order. In some areas such as mathematics and the science, in general more accurate measurements may be obtained (Though one ought to be weary about whether scientific and mathematical "progress" ultimately arises <u>from</u> such measurements). In other areas, located at the end of the spectrum, one finds the Humanities in which what is valid is not so easily identifiable. Maybe what is truly valid in the humanities is self-understanding; being able to move beyond self-deception through a constant questioning of the practices in which we engage ourselves and others on a daily basis. One such practice is that of testing each other.

But, where in this spectrum lies ESL teaching? That is a difficult call. Discrete grammar points are easily measurable. Learning to be bilingual, which implies strategic competence, discourse competence and communicative competence, represents a greater assessment challenge. If indeed our concern is with communicative competence, then there arises a radical tension

between validity and practicality. To this we shall in return in our next question. Let us just say that Brown's conclusion is that in the case of Standardized exams such as the TOEFL, the lack of any validity as regards oral proficiency is actually made up by other factors. ("what such proficiency test lack in validity, they gain in practicality and reliability" (pg. 387)). Nonetheless, from the idealistic approach, this seems to sound something like "the end justifies the means".

And besides, one need to seriously consider the following testing issues; how is one to validly test strategic competence? Should one give points for "avoidance" strategies in an exam? Or how is one to test for cultural understanding, if as we have argued, language is a life form and understanding a culture is in great measure understanding the culture from which is springs? Just to mention a few of the difficult issues we have to try to grapple with. Be this as it may, validity has a threefold nature consisting of content, face and construct validity. Let us look at them in greater detail.

3a. content validity: The first sub-division of validity regards content. A test's content ought to be the content for which the student has been actually trained. A written examination for a paramedic passing his CPR is quite invalid. You have to stick your mouth into the mouth of your plastic doll. You cannot just rewrite your manual, for in an emergency you might turn to look for a pen. Similarly, our students in EFL environments are so used to writing their responses, that one needs to frequently tell them to try to answer without writing down anything. But more poignantly; let us say that during your classes you emphasize stress and rhythm over other linguistic features such as exact pronunciation. If at the end of the month your exam includes none of these features, then your students might justly revolt. They might say: "Well if my continuing in the institution depends on passing a given exam, well at least the course should in great measure aid us in passing it rather than not." Unfortunately most exams do not actually tap into these oral skills, and therefore much content validity fails in terms of how we, for the most part, really test our students. If the communicative approach is to be taken seriously, one has to seriously consider innovative modes of testing which will indeed focus on elements such as stress and rhythm; so difficult to acquire by our students whose languages are for the most part syllable-timed. (See *Teaching American English Pronunciation* Unit 6 "Connected Speech").

3b. face validity: Through out these assignments we have spoken of a radical turn in pedagogical relations between teachers and students. However, once we get to the "real stuff", namely, to

testing, it seems that the learner disappears. It is only with face validity that the learner is reintroduced into the testing apparatus. It seems all to obvious that, for instance, if I have given a course on Aristotle's *Ethics*, then surely face validity is guaranteed by focusing on a passage of the text which might have even been analyzed in one of the classes. Or in the ESL environment. Surely, if you are hired to prepare students specifically for the TOEFL exam, then you must not waste time on preparing students for interviews, no matter how much it hurts you to know that the communicative approach is closer to "the" truth, than mere exam-preparing courses. In contrast, if you are preparing for the Michigan Exam, then for sure prepare them for the oral interview to be presented! And yet, one would often hear students complain about how what they were tested on had little or no relation to what "in the face of it" had been done in the actual classroom by their teachers.

3c. <u>Construct validity</u>: This is, for me, the most problematic of all the three elements guaranteeing validity in an exam. According to Brown, in order to see if an exam actually has this type of validity one need ask: "Does his test actually tap into the theoretical construct as it has been defined?" But this is quite difficult to understand. For it is not easy to see how content validity and construct validity can be so neatly separated. It is as if one could separate the house from its dwellers. Take out these specific dwellers, and that is not the same house at all. Or worse yet, take out the dwellers and that simply is not a house, though it might have four walls and a bedroom and a bathroom. That is precisely what is disturbing about documentaries filmed in the post-war Serbia and Croatia. One sees thousands upon thousands of empty walls, but no houses or homes.

  In other words, Brown seems to be sympathetic to this dichotomy. As we have seen, for him the prime example of the necessity of such a separation is the presentation of the TOEFL exam. The TOEFL is lacking in content validity, but its construction is such that this need not make us revolt against it. For Brown: "standardized tests designed to be given to large number of students typically suffer from poor content validity but are redeemed though their construct validation" (pg. 389). We must concede realistically: "the omission of oral content from the TOEFL has been accepted as a necessity in the professional Community" (pg. 389). An extremely paradoxical and troubling statement, to say the least.

  "But, what is so troubling about this?", you might ask. And I shall try to answer; but bear in mind, reflecting on what is taken for granted is not an easy affair. I find it difficult to even

articulate what is so wrong here. But I will try. One of my most remembered students in an advanced course on writing had the some of the worst writing skills I have seen for an advanced student. How had he learned most of his English? Well he was a musical kind of guy. He had even sung for bands and actually gave an astonishing presentation of the relation between Heavy Metal and Classical music. Would he pass the TOEFL? I do not know for sure, but my bet is that he wouldn't. And yet his pronunciation and rhythm were absolutely beautiful. It would be odd to have his skills be set aside, so that construct validity can win over the day. As I have argued, there may be too many Jasons who have everything to lose from our testing mania.

But more fundamentally still, from an idealistic perspective on testing, the crucial dilemma here is that the consideration of an exam's having construct validity invokes a complex theoretical construct itself. Remember what Brown has said "Does his test actually tap into the theoretical construct as it has been defined?" How to challenge an exam if it sincerely meets its theoretical criteria? Does this criteria not make exams avoid fundamental questioning in the sense that tests will always be devised not out of the blue, but rather following strict theoretical concerns. Take as an example a question form the exam for US Citizenship in the USA:

10. What is the date of Independence Day?

What is the theoretical construct behind such a question? I believe it is that citizens of a country ought to share a minimal history to which they seem themselves as belonging to. But I assure you Black Americas, Chicanos and Native Americans would perhaps have quite different responses if they could, in earnest, answer beyond the construct validity which is required of them. This is why Serbs and Croats cannot get along as well; they cannot share a sense of belonging to the same historical narrative. This is why they destroyed the famous Mostar bridge.

What then would, in the end, make up a good exam? For Brown taking into consideration all of the above mentioned aspects will generate a decent, if not good, examination: " If in your language teaching you can attend to the practicality, reliability and validity, …then you are well on the way to making accurate judgments about competence of the learner with whom you are working'" (pg. 389). But what we have tried to argue is that, although Brown's analytical separations aid us in clarifying the constitutive aspects of an exam, they do not in and of

themselves provide the clue to answering difficult questions regarding the relation between these sometimes conflicting elements. A move towards more alternative type of examinations, which we will consider below, may perhaps create less tension between some of the elements making up an "ideal" exam. Moreover, it is by moving towards alternative testing procedures that both the idealistic and realistic approaches may have a say in the matter. For now, we turn once again to one of the conflicts generated by Brown's triad, the tension between content and construct validity.

QUESTION 2

   We have already said much about the existing tension between content and construct validity. What more could be said? Perhaps besides questioning the role of oral skills, or better the lack thereof in the TOEFL exam, one can also point to the issue of its section on writing. Why is it invalid in terms of content validity? From an idealistic approach, the answer is clear. Precisely because, as we argued in our previous assignment, writing is not a 30 minute affair. Great emphasis was placed there on the need to focus more on the process than on the product of writing. By doing so, learners have a greater chance of actually becoming better writers in the long run. This would involve coming to slowly understand activates such as drafting, editing and revising; all of which require a patient capacity for self-correction. But the TOEFL exam emphasizes instead the product itself. All the talk about learner-centered instruction seems, once again, to come to a standstill when faced with the reality of passing or not a test which will enable you to undertake a career abroad. The realistic approach might answer, as long as no other exam is devised there seems to be no viable alternative. The paradox continues and dilemmas remain.

   But since we said a lot about this in the previous question, I would like to focus instead on the second part of this question, namely that no final and objective measure of validity exists and that instead validity ultimately goes back to subjective positions held by testers and theorists. Once again we need consider the topics under examination. For greater accuracy is expected of an engineer whose bridge might actually fall from an imprecision, than from an artist for whom an excess of accuracy destroys the rise of authentic creativity. As Aristotle argued more than 2000 years ago: "for it is the mark of a trained mind never to expect more precision in the treatment of any subject than the nature of that subject permits." (*Nichomachean Ethics*, 1988, Middlesex, Penguin Books, 1094b20). Ethics for Aristotle is prone to particularity, to nuances

and grayish areas which cannot be given rigid rules to follow; though this does not at all mean that anything goes. Aristotle best prepares teachers to the unexpected nature of our learners. Besides, in philosophy the issue of objective validity has been radically questioned as seeking for the humanities a model taken from the measuring and measurable hard sciences. No case is more famous indeed than that of the role of psychoanalysis within psychology. (See for example Paul Ricouer's famous essay: "The question of proof in Freud's writings")

Or in the case of ESL. We already mentioned the fact that objectivity is lacking in the sense that the same exam will be perceived differently by multiple correctors. You give it a A, I give it an A-; you care too much for how long it is, I do not; you care that it is absolutely analytical and argumentative; I like rhetorical effects and jokes so that writing ceases to be a boring affair. And we could go on and on. However, as we argued above, through experienced teachers and peer-corrections such tendencies may be made to vary without exaggeration.

But more importantly than this is the following issue. Tests are generated by paradigms. So if you are a audiolingual teacher you will prepare audio-lingual type of exams. To exaggerate, if you can keep up 45 continuous drillings, you pass. But the audiolingual tester does not, cannot even see, the relevance of contextual presentation. As we said, the TOEFL exam is itself blind to oral proficiency. A communicative exam itself may in the end be blind to the grammatical complexity of certain structures. In other words, as we argued in your Third Assignment, we teachers ought to be careful for we are for the most part informed by paradigms. In other words, the kind of paradigm which you defend –consciously or not--- defines to a great extent what we test for, in fact whether we test at all or not. This is the main reason for not leaving the idealistic approach to assessment on the side. It constantly reminds us of the dangers of over-evaluating.

Perhaps the best examples are provided by science. The very rise of the experimental method itself created a new kind of reality, a measurable reality. I cannot go into details here, but Descartes thinks he "proves", in his *Discourse on Method*, the philosophical underpinnings of this new type of relation to the world. And that all this approach may be quite blind is made more explicit by the fact that some have argued (Heidegger, Nietzsche) that this forcing reality to become measurable has led to an abuse of nature. Our current ecological crisis seems to reveal that the desire to measure can end up by simply destroying that which is being measured. Why in the end is nature measured? Well to be used. And a certain blindness remains as regards the demands that nature may place on us. Once again. as I have pointed out throughout all these assignments, we need listen to Heidegger's considerations as ESL teachers. Language is the

house of Being, not simply <u>our</u> house. Some things in nature, ourselves included, are truly better off when we decide that measuring them may violate them.

   So what does this have to do with testing in an ESL environment? Well that the all too realistic presuppositions of Brown and others, need be constantly challenged by idealistic perspectives such as those originating from the humanities in which testing really serves no purpose other than testing who you are, rather than what or how much you can do. What may be revealed is a desire to measure that escapes the most important aspects of education itself. A desire which may be seen to arise from a need to control nature, our bodies or our students. As against this desire for objectifying validation, one need recall the uniqueness of each of our learners. Maybe this is what Aristotle is trying to get at when he writes:

> "Moreover, individual tuition, like individual treatment in medicine, is actually superior to the public sort. For example, as a general rule rest and fasting are beneficial in a case of fever, but not perhaps, for a particular patient; and presumably a boxing instructor does not make all his pupils adopt the same manner of fighting. *It would seem, that particular cases receive more accurate treatment when individual attention is given, because then each person is more likely to get what suits him*" (Book X 1180b6-12; my emphasis)

Standardized testing, which fails for the most part to bridge the gap between content validity and construct validity, and which besides seeks a certain numerical objectivity as its overall goal, cannot quite deal with the uniqueness of individual learners. No wonder that Brown bids us, in Chapter 22, move from traditional testing to alternative testing (as much as we need to move from traditional medicine to alternative ones). Or better yet, instead of simply moving from one <u>to</u> the other, to   ascertain a certain bilingualism which acknowledges the presence of both approaches and can thus generate a healthy and changing disbalance between them by taking into consideration concrete realities and personal circumstances.

QUESTION 3

  We have previously made it clear that our modern industrialized and post-industrialized societies are caught up in educational dilemmas unknown to the past. One of the specific dilemmas in the ESL area is that of giving large scale communicative tests while maintaining practicality, that is to say, "the feasibility of scoring thousands of tests relatively quickly and cheaply". (pg. 394). If setting-up the infrastructure for the TOEFL exam, even without the oral proficiency section, seems to be quite demanding administratively speaking, it is mind boggling to even imagine how it could be done with the oral proficiency section in place. But if the

TOEFL might not change as swiftly as we would desire, we as teachers can move towards filling these "inevitable" lacks in our own classrooms. In order to do so, one would follow Bachman's emphasis of the four major elements which would be taken up by exams interested in tapping communicative competence sources. At last four considerations would become paramount: 1. use of realistic information gaps, 2. development of task dependency, 3. integration of new tasks and discourse within the exam, and finally, 4. assessment of a much broader range of communicative abilities and competences.

But perhaps the most interesting example of trying to provide an alternative is that of Swain. To picture the exam one could think of it as exemplifying the famous grid provided by the game tic-tac-toe. The "tic" tests for grammatical competence; the "tac" for discourse competence; and the "toe" for sociolinguistic competence. (Note, however, that strategic competence is nowhere to be found, for coming up with ways to assess it is quite difficult indeed!) And as in tic-tac-toe a three leveled assessment occurs within each column. Tics, tacs and toes include a triad of tasks: oral, written and multiple choice exercises. (Brown pg. 397)

But does this innovative presentation lead us out of the dilemma? According to Brown it is absolutely impossible to replace the much more inexpensive TOEFL exam (Brown, pg. 389) (though if your earnings are in pesos, or other developing country currency, then "inexpensive" seems not to be an adequate word). If this is so, the dilemma still seems to haunt us. Or does it? Perhaps by moving in the direction of alternative assessment in our classrooms, students might understand that although there are set in place some exams which must be presented, the actual skills which they will require to pass it and, moreover, to really interact in the given target culture will have been provided within the classroom itself.

But what then could be done with Swain's exam? One would need much information on what these testing mechanism actually, realistically, require in terms of infrastructure. The defenders of the realistic approach to test-taking have much to teach us here. However, at the very least it could be considered as an optional exam to the TOEFL. If students have the money to present it, it should be possible for them to do so. And if coming from a developing country, specially for postgraduate students, the Universities could pay part of its costs. Nonetheless, moving beyond the dilemma requires much more than particular and punctual changes in individual exams. It requires a paradigm shift altogether, a new way of understanding what assessing ourselves and our learners is all about. These changes can best be understood by

focusing on alternative modes of evaluation. By doing so the gap between the idealistic and the realistic approaches to testing can be lessened, rather than dangerously widened in the long run .

CHAPTER 22: LANGUAGE ASSESSMENT: PRACTICAL CLASSROOM APPLICATION

QUESTION 1

In order to confront the dilemmas we have signaled to in our previous answers, it is relevant to review the meaning of the different testing modes. In this sense it is helpful to enter the terminological maze of assessment.

1. informal testing: One thing is a party which requires a tie, quite another one in which blue jeans are acceptable. Informal testers prefer to wear jeans. To test informally relates to all incidental and unplanned evaluations which are not carried out for the purpose of recording measurable or statistical results.(Brown pg. 402) This seems quite complex. However, the best example of informal testing is our everyday interaction with students. Whatever production we comment upon, on whichever of the skills, is liable to become part of our informal assessment. Students may not even be aware that we are assessing them. Continuous assessment becomes a priority from this stance. Constant feedback and the ensuing dialogue between teacher and learner generates the conditions for a friendlier atmosphere towards pleasurable learning. Informal testing is, to use computer terminology, much more "user friendly". It is quite different to say "I will test you tomorrow", or "WE will embark in continuous assessment". In this regard words do make a big difference. Just as wearing a jean or a tie generates a different stance between the participants at a party.

2. formative testing: If teaching has anything to do with creating "form" -------that is to say, giving form to what at the start may appear quite amorphous---- then formative testing becomes central to teaching. I recently had the, once in a life time opportunity, of looking at Degas' Dancers in the Art Gallery of Ontario. Degas produced most of them by using rudimentary elements and clay. Only later were they finally set in bronze. Giving form to what is truly quite unappealing, sticky and brownish clay requires an artful character. To generate from this opaque and bland material, tri-dimensional figures of dancing women and girls who nonetheless remain still, this requires great genius. Clay, a simple piece of matter, is given diverse forms by able

artists who recognize the strengths and weaknesses of the particular clay found in their regions. Some clay is reddish. Some brownish. Some clay is finer, some is rougher. And yet from each, beautifully finished artifacts ---vases. jugs, ceremonial receptacles and others ---- can be created.

  So what does this have to do with the matter at hand, namely, "formative testing"? Well, that this kind of testing seeks to form, rather than solely to measure. You may measure, weigh and analyze clay, and yet your clay figures may end up being undistinguishable from mud. Standardized testing in contrast, intent on measuring, may in the end not be capable of generating long-lasting forms. Or if it does, just because it is standardized it may even end up producing indistinguishable pieces of less than exceptional art. Of course, our students are not clay, nor are we potters who make them; but what we can provide through formative testing is one of the means for having students engage themselves as if they we the most important clay around, namely, their own clay. Giving shape to yourself over and against the tenacious impulse to conform to prefabricated molds is quite a daunting task. Some of us turn out to be quite formless, other of us too rigidly formed. Some of us cannot stand up, some of us seem to shatter easily under pressure. Testing has a similar effect, it can shape, it can shatter.

  Perhaps one way of providing greater flexibility to the educational process is by changing our idea of what assessment is. Instead on focusing on the final product (the midterm, the final exam, the test for admittance, the final report, these assignments), one could focus on the process of assessment generation itself. Although the product ----as we saw in teaching the writing skill----- does not cease to be of great importance (for how to evaluate a draft and expect students to forget about their final essays), learning about what goes on in the making, and about what one ought to avoid in this making of good/excellent products, becomes essential. Too much water, and the clay will not stand. Too little, and it cannot be handled with ease. Formative testing or assessment involves an ongoing process in which constant feedback moves back and forth from teacher to student. Specially in a language classroom, where most of the interaction is not realistic but only an approach to the to the real out-of-classroom environment, formative assessment looks at on-going output in a space where students can feel more at ease. And yet, ironically, this type of assessment may generate another kind of "pressure", one which is self-generating. The kind of tension we feel when we are thrust forward by a question to which we seek some, at least, partial answers. The kind of tension we feel when we have a dim sense of our not truly knowing who we are. In sum, if in the area of formative testing continuous assessments

plays a core role, then it is important to get clear of what continuously testing our students might mean. In this respect:

> "continuous assessment should be regarded as an integral part of your teaching and your student learning. Above all, it should be designed and administered so that it forms a pleasant component of your teaching program. Often students will be unaware of any kind of assessment taking place since the whole situation will be informal and relaxed. If continuous assessment is treated as a formal means of measurement in any way at all, its value will be largely lost" (pg. 112 Continuous Assessment in Language Testing Terminology. Photocopies from other training courses; unfortunately I do not have the reference.)

Taking on seriously the pull we feel from the scientific paradigm in our societies, it seems as though what is not measured cannot be assessed. But, again, this is not the case for many of our daily activities, and therefore we should feel more at ease when searching for alternative modes of testing. These may even begin to acknowledge that some subjective factors always enter the assessment process.

 What would formative testing and continuous testing look like? It would be a little messier than normal testing. This is so because the teacher is assessing even while students are interacting, even while the actual teachings going on. Continuous assessment proceeds rather through notes, through day to day formats. By proceeding this way it will provide, or at least one would hope so, a more authentic and holistic view of the majority of learners. I will provide one of my own messy attempts at this kind of assessment; it is one of the formats used for my advances course entitled "Grammar Though Composition":

It may seem amorphous, messy and with badly measured lines. However, out of this less-than-perfect product, a more realistic consideration of each student may arise. To continue with the clay metaphor; the above provided sheet of paper was the clay out of which I tried to provide a fair and authentic assessment of each of my students' production. The final product itself was an electronic magazine   which brought us to understand that in the duo "formative testing", the much more important bit is the "formative " one.

3. <u>formal testing</u>: Rarely do we hear either positive things or a sustained argumentative defense of this type of testing. According to Brown, formal testing requires much more of a tie. You've had a chance to play around with the clay. You have got 30 minutes to produce a vase. In other words we provide specific exercises designed to get students to let us know how much they have learned about a given topic, normally within a relatively short time limit. Moreover, formal testing is systemic and preplanned. We have said much about these traditional testing devices above. Their paradoxical nature is once again stated by Brown when he writes:

> "From pop quizzes to final exams to standardized entrance exams, tests are crucial milestones in the journey to success. It is unfortunate that learners all too often view tests as dark clouds hanging over their heads, upsetting them with lightning volts of anxiety as they anticipate a hail of questions, they can't answer and, worst of all, a flood of disappointment if they don't make the grade" (pg. 401)

The paradox arises for, on the one hand, we must acknowledge following the realistic approach that they "are crucial milestones in the journey to success'. And yet in Brown's own wording we must realize that it is quite "unfortunate" that they produce some nasty secondary effects. The paradox could be put in medical terms. If you have arthritis you might have been give a drug called "methotrexate". The list of secondary effects is so long that it takes quite some time to read. Would <u>you</u> take it? However, traditional perspectives will let you know from the start that they are crucial for success, however unfortunate the effects may be. In other words, although "formative" and "formal" stem from the same root, what they mean by "form" is two very different things.

   From an idealistic perspective Brown seems to be, at least partially, blind. And from a realistic perspective, this realization requires a move towards a combination of formal and informal mechanism of assessment. To these combinations we shall return below.

4. <u>summative testing</u>: In contrast to formative testing, where the overall process is what we pay attention to when assessing our students performance, the main concern here is the final product itself. You had your hours of clay training, now you have got to produce one product. You have to summarize what you know in one specific and concrete example. How else could I know that you can <u>actually</u> makes vases? Or as Brown puts it: "By nature such assessments tend to focus on products of learning: observable performance that may be evaluated somewhat independent of the process the student has traversed to reach the end product" (pg. 403). The end product is what counts. So your clay was not of the best quality; well, that is just quite unlucky. So your clay teacher was not of the greatest quality; well, that is truly unfortunate. So you had a terrible headache last night because of some difficult situation or other; well you should have thought about it before taking up this course (you knew full well the final exam was on such and such a date).

Since summative testing is the type to which we have become more accustomed ---we moderns that is--- I will focus on the issue of how these tests might provide constructive feedback on students. The best example is the following. At the CCA we had 1 month courses for 2 hours everyday (now it has changed a bit). The summative CCA examinations ---which have a very complex history that gave me multiple examinations formats to learn from (i.e., periods of non-examination, of pass or fail, of total continuous assessment, of very regular exams)----- for most courses were on the second to last day of the course, namely, day 18. The tension in the institution would reach blackout proportions. Tense teachers, tense administrators, tense secretaries, and above all tense students. After the release of tension by all of those participating in the process, day 19 was used to slowly and carefully go over the summative exam with students. (we shall return to pre, during and after the test recommendations on page 409 in what follows.) In this sense the summative product itself was seen as an aid in the learning process.

Unfortunately, these tactics and strategies for test taking, important as they are, do not deal with the core problem of summative exams. What is the central concern? Well, the fact that people **fail** them. Most summative tests are either to be passed or failed, you cannot have both, you cannot have the cake and eat it too. You learn, you pass; you cannot learn and not pass. And failing is something which is not easy to talk about or digest. A teacher might react, as we usually have had to, "Well you failed, any questions?" But that kind of misses the point. Personally speaking, and I hope I am not the only one, in countries like Colombia where a harsh

reality makes self-esteem reach quite low points (you cannot imagine how Colombians are treated at airports around the world), failing is equivalent, for many, to being less. Looking back and recalling students crying because they had failed an exam is something which I find all too shocking nowadays.

More concretely. At the CCA the passing grade for some of the years I evaluated was 3.5 . You know as well as I that for many examinations there is usually a predictable kind of curve, few excellent, some/many good, some/many borderlines, few really bad failures. The typical case over which discussions were done, again and again on TDTS sessions, was what to do with a person with a 3.4 or 3.3. (Although we as teachers provided an oral grade worth 30 percent, the rule said –at the time--- that if you failed the exam you failed the course). The teacher's energy was spent on deciding the future of his or her students. To make a long story short, the final day of correction this is what I did after carefully correcting the summative test. I congratulated those good results which always seem to make a teacher proud by giving them a public hearing while reminding them how much of what we had done never actually came into consideration in the exam. As for the 3.3. or 3.4 who had good speaking abilities, I would set them aside. I told them that it was up to them to pass or fail; whatever they decided I would "arrange". Not all had this chance because my "formative assessment" told me that, honestly. it was best for some to focus on certain issues rather than end the program and be unable to speak English at even the intermediate level! This exercise was interesting. Many who certainly did not deserve to pass passed, and many whom one would have never failed, failed themselves.

In other words, what is most questionable about formal and summative testing is that it places teachers in troubling situations involving heart felt cries, threats and every once in a while some swearing. It does so at the expense of making of the learning process instead an exciting and enticing cooperative project. The only thing that can be said here is that months later I would keep in touch with many of my students who in general were incredibly nice to me. They would tell me that they did not regret repeating a given course. We teachers have great power and most of us have from time to time abused this power in learning to become better teachers ourselves.

QUESTION 2

Moving beyond the testing paradoxes to which we have alluded may mean not simply doing away with assessment altogether, but rather attempting to redefine and alter what assessment is

all about. Assessment need not hamper growth, but instead may actually promote growth itself. For nothing is more disillusioning for a learner than to receive little or no feedback on her work. This is the case for most of our best students who see no challenge in our commentaries. And besides, if no relevant feedback is provided, then "why do we need teachers anyway", a student might come to think. As our previous discussion has focused more on theoretical issues, here I will, following Brown, provide examples of what alternative testing might look like. I have already provided the example of continuous assessment for the advanced writing course entitled "Grammar through composition". As you might recall that was altogether a very messy affair. But before going ahead and providing examples one could first of all mention some of the main characteristics of this type of assessment: it is continuous, it seeks long-term objectives, it is untimed, it has a free-response format, it seeks to be contextualized, communicative proficiency is central to its appearance, it is task- oriented, it is formative rather than summative, it seeks interactive feedback, it fosters open-ended and creative answers, it is process oriented, it generates interactive performance, and finally, it sees assessment –something almost unbelievable--- as generating intrinsic motivation. Now to the examples.

1. portfolios: As part of our 50 hour training per year at the CCA we had to present formative portfolios at the end of the year. Most teachers did not much get into the idea; which is quite odd because some of the training session were about informal assessment itself! I myself presented at least two. A portfolio is a much more authentic type of assessment for it takes into consideration a longer period of time. Portfolios share the nature of a story; they provide the author and the reader with multiple aspects which might have changed though time. It is composed of a multiplicity of elements which generate a stronger and more accurate picture of who one is in terms of one's actual working decisions. When it is finished you are astonished, you ask yourself; "Have I done all those things? But, when?" I cannot provide the complete portfolios for they are quite lengthy, but from them I have two examples to provide. In my end-of-year portfolio for 1997 we had to answer several interesting questions (some of the responses are quite astonishing to me, as they argue for very similar points as those I have argued for in these assignments). But one of the questions was precisely to try to provide a self generated definition of what a portfolio actually is. This is what I came up with 6 years ago:

THE CONCEPT OF A PORTFOLIO

A portfolio can be likened to an artist's exposition; it is somewhat like the ones we have the opportunity to see at the Colombo itself. Fortunately though, our portfolio is not done *post-mortem*. The comparison can be a little presumptuous, yet still, when you see an artist's work you have a chance to see the important transformations in her styles. Such I believe could be a metaphor for a portfolio. By looking at the different portfolios one could start to comprehend the differences, not only in an individual teacher's work, but likewise in the diverse backgrounds and teaching styles among the multiple teachers. And since you yourself participate in the exposition, then clearly the other portfolios present themselves as mirrors through which you can really get to understand your own teaching style in a clearer way. You come to realize that some works impress you more, and have a greater impact on you, given their characteristics; characteristics which you might share or at least learn from."

Most probably I would not write the same things today. But what is clear is that portfolios create an atmosphere in which self-understanding becomes much more central than a grading scale. Another example is that of my 1999 portfolio. I will simply present the content table to have a feel for the diversity and complexity of its components:

"**PORTFOLIO: CONTENTS**

A. EVALUATION SURVEYS BY THE COLOMBO

B. READING GUIDES FOR THE TDTS GATHERINGS

C. REPORT BY ADs AS WELL AS REPORTS ON CLASSES OBSERVED

D. EXERCISES USED FOR THE ADVANCED LITERATURE COURSE

E. EXERCISES AND FORMATS USED FOR THE ADVANCED COURSE ON ENJOYING GRAMMAR.

F. IN-CLASS STUDENT SURVEYS"

This multiplicity of factors, which includes multiple feedbacks formats both from the institution and students, cannot but provide a more accurate picture of how to assess an individual than the picture traditional testing could ever do. As to the nature of this institutional feedback, let us just say that I myself was visited frequently by different teachers in my classrooms. The formats used as "Official Observation Checklists" could be quite complex. I have provide the first page of

one, so that you can have a feel for the seriousness of the CCA as regarded its developmental tools

CENTRO COLOMBO-AMERICANO
Teacher Observation Checklist

Teacher: Andrés Melo Course: Enjoying 6. Hour: 5-7
Visitor's Name: Nidia Cortés ☐ AD ☒ Supervisor ☐ Peer Date: 09-23-99
Activity in Progress: Untrue Conditionals Type of Visit: official

T HAS TAUGHT THIS COURSE
At least 3 times before .... ☐
Only 1 or 2 times before .... ☐
First time .... ☒

1. OBJECTIVES Y N
T had clear objectives .... ☒ ☐
T checked that objectives were reached .... ☒ ☐
Objectives were reached .... ☒ ☐

2. WARM-UP Y N
Warm-up was used .... ☒ ☐

To review ☐   To get Ss ready for next topic ☐
To motivate Ss ☒   As icebreakers ☐
To recycle ☐   As pre-activity for role play ☐
For integration ☐   Other ☐

3. PRESENTATION
(Structural/Functional/Lexical) Y N
☐ ☐ ☐
Observed .... ☒ ☐
T built up on previously learned material ...... ☒ ☐
Done with books:
Closed .... ☒   Open .... ☐
Done in context .... ☒ ☐
Done: Inductively ..... ☒   Deductively .... ☐
A combination of both .... ☐

T provided examples:
Contextualized .. ☒ Real life .. ☒ Simulated .. ☐
T involved Ss in presentations by eliciting:
Questions .... ☒   Examples .... ☒
Spoken responses .... ☒   Physical Responses ..... ☐
Contexts .... ☐   Definitions .... ☐
Other: _____

Presentation was:
Clear .... ☒   Concise .... ☒
Complete .... ☒   Graded .... ☒
There was drilling after each step .... ☒ ☐
T checked understanding:
By asking for examples .... ☒ ☐
By asking concept questions .... ☒ ☐
By setting a task .... ☒ ☐
Other: _____

4. METALANGUAGE
Appropriate .... ☒
Above Ss' level .... ☐
Below Ss' level .... ☐

VOCABULARY Y N
Presentation of vocabulary was observed ☒ ☐
Relevant vocabulary was presented by:
Using gestures .... ☐ Translating .... ☒
Using visuals .... ☒ Giving Contextualized examples ☒
Writing words on board . ☐ Defining .... ☐
Asking Ss for definitions . ☐ Drawing .... ☐
Other _____

5. TEACHING AIDS Y N
Used .... ☒ ☐
Used for:
Presentation ... ☐   Warm-up .... ☐
Practice .... ☒   Wrap-up .... ☐
Use of:
Realia .... ☐   Workbook .... ☐
Worksheets ..... ☐   Book .... ☒
Tape .... ☒   Blackboard .... ☒
Drawings .... ☐   Pictures .... ☐
Other _____
Used effectively .... ☐ ☐

6. PRACTICE Y N
Practice was appropriate for objectives .... ☒ ☐
Practice was contextualized .... ☒ ☐
T set realistic time limit with Ss .... ☒ ☐
T recycled when possible .... ☒ ☐
There was T-S involvement .... ☒ ☐
There was S-S involvement .... ☒ ☐
Logical sequence of activities .... ☒ ☐
Appropriate seating arrangement .... ☒ ☐
Variety of activities (skills, types, set up, etc.) ☒ ☐

7. TRANSITIONS Y N
Used between activities .... ☒ ☐
Used between topics .... ☐ ☐

8. INSTRUCTIONS Y N
Instructions were complete .... ☒ ☐
Instructions were concise .... ☒ ☐
Instructions were clear .... ☒ ☐
T provided a model .... ☒ ☐
T checked Ss' understanding .... ☒ ☐

In this particular case I was visited by Nidia; she is one of the most brilliant EFL teachers I have known.

But portfolios are quite shocking in another sense. To see how accurate a portfolio can be, I will tell you a secret. As I love to sketch whenever I can, many of my TDTS session papers (particular the one on evaluation!) was covered with sketches which I cut in order to further develop on my own. Months later we were asked by our director to actually include the formats themselves in our portfolio. I had to check my "Sketches Folder" in order to retrieve the cut-outs so that they could be pasted once again into the evaluation sheets themselves. So when looking at my portfolio you would find a lot of sketches all over the place! If they had not known me a little

better, I might have gotten into trouble. But seriously speaking, this was in fact MY Portfolio, one which includes in any piece of paper small drawings from which greater and more developed ones can actually come about. Alternative testing seems itself to place more importance on the sketches rather than on the actual final product.

 2. <u>story telling</u>: Another possibility for alternative examination is that of storytelling. Almost all cultures have as part of their identity stories which configure the realm of the imaginary for them. (See Alasdair MacIntyre, *After Virtue*). You might have a story about "Peter Pan"; we have stories like that of "La pata sola", a woman with only one leg. How could this type of alternative assessment be carried out? It mostly depends on the skill being emphasized; but in the case of writing, students may be provided with the following figures:

What does the test consist in? First of all, in cutting up the small squares (physical response), arranging them in whatever order they please (creativity), and finally generating a storyline by using discourse connectors and including a concluding statement (discourse competence). This particular set of figures generated incredibly creative stories which never ceased to surprise me. Dinosaurs, spaceships and aliens; "well is all <u>that</u> realistic, truly communicative?", you might ask.   Well not really, but if learning language requires being imaginative, and if alternative assessment is to intrinsically motivating, then surely these factors far outweigh others. To sum up:

"Storytelling is traditional in most cultures. We can tap into that tradition for a very portable resource and a convenient and flexible technique for teaching any phase of a grammar lesson. A story provides a realistic context for presenting grammar points and holds and focuses student's attention in a way that no other technique can. Although some teachers are better at telling stories than others, almost any of us can tell stories with energy and interest. Students naturally like to listen to stories, and most remembered long after the lesson is over" (Chapter Four, *Techniques and Resources in teaching Grammar , O.U.P* 1988)

In this respect often, when meeting someone for the first time, we will say: "what is the story of your life." That is to say, we seek a narrative to make ourselves understood to ourselves and hopefully to others who share similar stories and similar puzzles.

3. Sketches: As part of the resources from Heinemann English Language Teaching, Doug Case and Ken Wilson have created   *English Sketches*. Not only are they fun, challenging and demanding, but they are also photocopiable. They likewise require great levels of creativity and invite students to tap into solution-oriented thought modes rather than only into memorization and repetition skills. The sketches themselves are not too complex for intermediate students (they have created different sketches for different levels). Some, as myself, cannot act and therefore truly admire those who can (unless one recognizes that teaching requires much acting). In the same way, some students might feel awkward in doing them. But my experience is that even those will finally loosen up and will truly thank you for the possibility of expressing themselves in ways they never thought possible. So that you have a feel for the kind of comic relations dealt with, I will provide a short excerpt from *The Ticket Inspector*:

"Steward: Coffee!
Passenger: No, thanks.

Waiter: Seats for dinner!
Passenger: NO, thanks.

Inspector: Tickets!
Passenger: No, thanks.
I: Pardon?
P: I don't want a ticket, thank you.
I: I'm not *selling* tickets , sir.
P: No?

I: No, I want to see *your* ticket.
P: Oh, I haven't got a ticket.
I: You haven't got a ticket?
P: No, I never buy a ticket.
I: Why not?
P: Well, they *are* very expensive, you know."

 By assessing elements such as rhythm, stress and intonation, oral proficiency can become the center of our continuous assessment rather than a mere appendix which will severely cripple our student's chances to interact with the target culture. The only thing we might fear from this particular exercise is that our students should perform them *verbatim* when boarding a VIA train!

4. <u>electronic magazines</u>. As I have mentioned in my other assignments, I created at least three electronic magazines for my advanced writing, grammar and literature courses. Although generating them requires much outside-the-classroom time, I do not know of a more motivating ways of assessing students by having them engage in the process of self-correction. Student do not really care too much for writing, as was made clear in our previous assignment. This is why it is quite challenging to get them to write at all. Particularly so in the case of literature, for perhaps the focus on too much realism within the ESL environment has hampered the need for students to understand that, although there survival techniques require realistic English, the <u>overall</u> objective should be much more than that. Reading and writing on literary topics seems to me an ideal goal, if in fact it is in reading and writing that we explore ourselves beyond what we taken for granted. (I have provided a copy of one such magazine to Benns International Schools)

5. <u>peer correction</u>: As I have been able to visit, during a few days, the classroom equivalent to this distance course, I have been able to see first-hand the constant use of peer correction techniques as part of an overall change of what assessment is all about. After any given presentation all those present write commentaries on what the presenter had to say on, for instance, a given grammatical topic issue. Reading them out loud generates a certain kind of public responsibility which can jumpstart, in most, a desire to prepare as much as possible. Another example of this type of peer assessment is the following:

Figure 4.11 Peer Feedback Form: Explaining a Process

Speaker's Name _____ Date _____

Your Name _____

Part 1: Circle the word Yes, Some or No to tell how you feel about the speaker's report

1. I understood what the speaker was talking about.	Yes	Some	No
2. The speaker describes how everything worked.	Yes	Some	No
3. The speaker explained in steps I could follow.	Yes	Some	No
4. I think I could do this myself now.	Yes	Some	No
5. The directions were clear.	Yes	Some	No

Part 2: Complete the following sentences.

6. I liked when the speaker _____

7. The speaker was good at _____

8. Maybe the speaker could _____

Adapted by ESL teacher M. Crossman from Hill and Ruptic (1994).
• ©Addison-Wesley. *Authentic Assessment for English Language Learners.* O'Malley/Valdez Pierce. This page may be reproduced for classroom use

160

(unknown title of source; the format is small in order for me to be able to send it through the Internet)

I myself provided feedback on multiple teachers at the CCA. One of the reports I treasure much for I must seriously say that many aspects of my EFL teaching skills are due to Edmundo, a colleague at the CCA.

6. <u>Reading exam with cassette and oral peer correction</u>

One of the kinds of assessments which can provide students with the possibility of better understanding the issues underlying stress and intonation, is that of taping live readings. Although at first students may feel awkward when being taped, once they hear their own voices and actually see how the stresses have been misplaced, they forget the initial awkwardness to which they were "submitted". But, most importantly, they can learn to tape themselves whenever and wherever they wish. Since most material comes with a corresponding cassette, they can actually compare their intonation and rhythm with a model from which they can learn much. Of course, it is not realistic to read out loud ----though it seems to be a lot of fun to do so (just ask

in a classroom if somebody wants to read out loud)--- but the aim will be to help learners transfer these skills into the overall communicative skills which they must develop.

7. <u>They are the teachers</u>. One kind of alternative assessment which I love, and I do not see mentioned all that often, is to have students becomes the graders of lower level students. By doing so they come to learn the difficulties in grading, assessing and correcting. The following exercise I did as part of an advanced writing course. I gave them stories which had been written by students from Course VI-B (there were at the time 12 basic levels at the CCA), and had <u>them</u> correct their intermediate peers by using both the writing symbols which I provided in my last assignment, and to give them a numerical overall grade. As an overall conclusion one could say that they had fun; it is fun to have power but not have to be in real situations where it has to be applied. But much more importantly, they came to realize how drastically they corrected their own. One would see, and let them know, how tough they would be if they were ever to truly correct students.
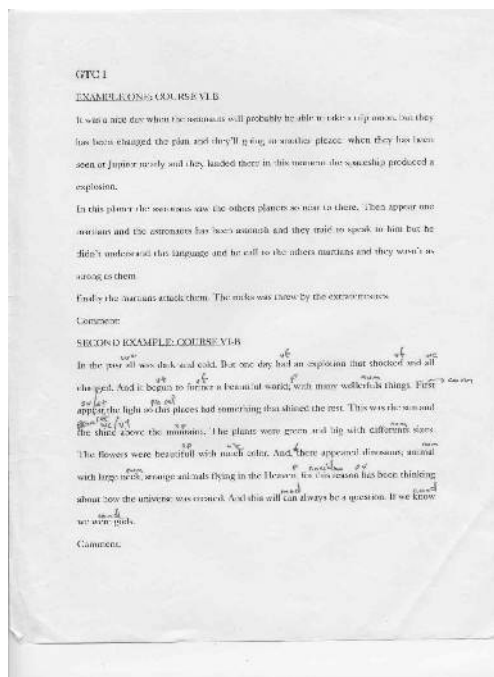
(the format is small in order for me to be able to send it through the Internet)

8. <u>self-correction:</u> The move from traditional assessments to alternative modes of evaluation is absolutely essential if considerations taken from the idealistic approach to testing can better correlate with realistic concerns as well. However, within the camp of alternative testing it is my

view that not all testing carries equal weight. By far the most important type of testing which should arise out of alternative testing is that of self-assessment. In an ideal world if all of us were the best self-assessors around, then teaching would end up focusing all its energy on what actually counts, puzzling creatively about issues which can generate a multiplicity of solutions to diverse thematic. If our move towards a more learner-centered education is to carry any weight, then assessment should itself become learner-centered. Self-assessment and autonomy seem to go hand in hand. Individuals capable of assessing who they, and where they stand on a given issue, are best equipped to generate comprehensible relations among us. In this case the following format for oral self-assessment could be of some aid:

to what they hear, whether it's taking notes, charting a route on a map, or answering questions (Ur 1984). Tasks should be designed to challenge the proficiency level(s) of your students without frustrating them.

Oral language assessment can take various forms depending on your purpose for assessment, students' level of language proficiency, and the purposes for which students use oral language in the classroom. Assessment tasks for oral language differ with regard to whether they call for the use of static relationships (such as in describing a picture or giving directions), dynamic relationships (telling a story or taking part in a role-play), or abstract relationships (giving an opinion) (Brown and Yule 1983). These relationships correspond to an increase in difficulty levels; that is, it is easier to describe a picture than to give an opinion in one's second language. You need to consider the purpose of the assessment, the format (individual, pairs, groups), students' level of proficiency, language functions used in daily classroom activities, and the level of student preparation needed for each assessment activity before choosing it. You also need to consider the difficulty of tasks with regard to both linguistic and cognitive load; the tasks should be developmentally appropriate and authentic. The important thing is to give students continued opportunities to engage in authentic oral language activities before using these same activities for assessment.

Figure 4.6 Self- Assessment of oral language

Name _____   Date _____

Check (3) the box that shows what you can do, Add comments.

What Can You do in English?	Difficulty Level				Comments
	Not very Well	Okay	Well	Very well	
1. I can ask questions in class.					
2. I can understand others when working in a group					
3. I can understand television shows					
4. I can speak with native speakers outside of school.					
5. I can talk on the phone.					
6 I can ask for an explanation.					

Adapted from Bachman and Palmer (1989).
• © Addison-Wesley. Authentic Assessment for English language Learners. O'Malley/Valdez Pierce. This page may be reproduced for classroom use.

155

(unknown title of source; the format is small in order for me to be able to send it through the Internet)
Andrés Melo Cousineau

And yet such a shift cannot be done from one day to another, for it requires not only a radical paradigm shift, but a transformation of our desire for recognition. It may even take a whole generation for some results to develop; perhaps because we teachers have for the most part grown accustomed to teaching by means of traditional assessment. Moreover, students themselves need learn how to assess themselves adequately, for they might end up just seeing how excellent they are! The Focus On Grammar Series is here helpful once again. This is so because at the end of several units it provides a "self-test" for which all answers are found at the end of each book. By having students test themselves, and of course aiding them whenever they might have concrete questions, they may begin to understand that, in the end, the teacher is to a great extent irrelevant in the learning process. What is relevant overall is that the learner (including the teacher) learn, not that the teacher teach.

  And we teachers can learn form this dynamic ourselves. For as teachers we may come to unknowingly create rigid frameworks around from which it is difficult to "escape". It is not easy for us teachers to accept criticism from our peers. It could be because we ourselves are used to seeing assessment as playing a negative role in our educational process. It could be because we are too caught up in being recognized. But if we could ever come to understand that another might better see our limitations, then perhaps we would learn to welcome another's words. These would be seen not as threatening, but would instead be welcomed. One such format for teacher self-assessments concerns how far we actually use authentic assessment:

Figure 2.2 Authentic Assessment Inventory for Goal Setting

Teacher _____ Date _____

Purpose: This inventory will help you establish goals for further development and use of authentic assessments in your classroom.

Directions: After reading each statement, circle the appropriate number in both columns to indicate (1) how you are using authentic assessment approaches, and (2) your ideal level of use.

1	2	3
Not at all	Somewhat	A great deal

To What Extent Do I:	Where I' am Now:	Where I Would Like to Be:
1. want to use authentic assessments?	1 2 3	1 2 3
2. want to use authentic assessments more effectively?	1 2 3	1 2 3
3. clearly define levels of student performance?	1 2 3	1 2 3
4. plan scoring rubrics before using assessments?	1 2 3	1 2 3
5. compare student performance to a standard?	1 2 3	1 2 3
6. inform students about scoring criteria before judging?	1 2 3	1 2 3
7. find authentic assessments difficult to use?	1 2 3	1 2 3
8. feel concern about the time required to use them?	1 2 3	1 2 3
9. talk to other teachers about authentic assessment?	1 2 3	1 2 3
10. share my assessment strategies with other teachers?	1 2 3	1 2 3
11. share teaching strategies for authentic assessments?	1 2 3	1 2 3
12. ask students to rate their own performance?	1 2 3	1 2 3
13. ask students to rate each others' performance?	1 2 3	1 2 3
14. give students feedback about their performance?	1 2 3	1 2 3
15. give parents feedback about their child's performance?	1 2 3	1 2 3

Use: Review the two right-hand columns to identify differences between where you are now and where you would like to be. Items for which there is a difference of one or two points can become target areas to establish goals for authentic assessment. Circle the statement for those areas.

Adapted from Stiggins (1992).

• © Addison-Wesley. Authentic Assessment for English Language Learners. O'Malley/Valdez Pierce.
   This page may be reproduced for classroom use.

120

(unknown title of source; the format is small in order for me to be able to send it through the Internet)

In order to conclude this section on alternative assessment, let me make a parallel with medicine. Traditional medicine focuses on an illness' diagnosis. Tests upon tests are required to reach a final diagnosis. The diagnosis is the end product which, once it has been given, determines the totality of a patient's life. "You have got cancer" eventually turns into "the cancer having you". In other words, the diagnosis makes continuous self-assessment on the part of the doctor and the patient, highly irrelevant. In some cases diagnosis are absolutely damaging for

they close most doors for other modes of medical understanding.. Listen to the following case which I have come to believe is not all that uncommon:

> "You won't believe what doctors did to me in Finland," she confided. I asked for details. "It took the a long time to make the diagnosis, many tests. Then finally, the head neurologist took me into his office and told me that I had multiple sclerosis. He let that sink in; then he went out of the room and returned with a wheelchair. This he told me to sit in. I said, "why should I sit in your wheelchair?. He said I was to buy a wheelchair and sit in it for an hour a day to "practice" for when I would be totally disabled. Can you imagine?" She related this tale with a healthy laugh….." (pg. 64 Andrew Weil, *Spontaneous healing*)

So what is the relation with our area? At least two fold. First of all, perhaps that too much testing in teaching –particularly traditional, but even maybe alternative---- can actually miss the point of learning. At the very least we hope we are not making our tests sicker, rather than healthier. Hopefully, we are not moving learners towards wheelchairs rather than helping them to walk on their own; or why not, even to dance on their own or with others such as Degas bid us do. But secondly, it is clear that changes in self-assessment are slow, difficult and in some cases even painful given human nature. If indeed ethical understating and medicine have important parallels ---as Socrates, Plato and Aristotle seemed to have thought--- then the general trend by people to prefer alternative testing on both counts seems to point towards newer forms of understanding ourselves. But such transitions come, most of the times, at a cost. We lose the security to which we have grown accustomed. And this requires acknowledging the fragile nature of such interim conditions. Not providing enough space, enough time for such general recoveries may even be worse than even beginning the process altogether. This is why I must here, once again, remind myself and others of Socrates' life long dedication -–—which ended in his death by one of the most "brilliant" cultures ever to have existed--- to a kind of self-assessment which philosophy can truly come to grips with. At his trial Socrates, a few days before his death, is said to have said:

> "if on the other hand I say that this happens to be the greatest good for a human being ---each day to make rational arguments about virtue and also about other matters concerning which you hear me carrying on dialogues and examining both myself and others: and the unexamined life is not worth living for a human being ---you will believe me still less when I say these things. But it is so, as I affirm, men ---though to persuade of it is not easy."

QUESTION 3

With regards to strategies for taking exams, consider this imaginary case. Standardized exams are allegedly set in place in order to keep a minimum of standards for the educational development of a given population, usually a national one. Know, let us suppose that if fact <u>all</u> our students became very good test-takers, and eventually were continuously graded as excellent in terms of these exams. Troubling questions would arise, as far as I am concerned, from within the idealist camp. For one, would we be satisfied with this? I mean, what more could you ask of education if tests actually show that students excel? Or more seriously, why would one really need teachers, instead of instructors or computers? Why not come up with complex electronic creatures who could coach students towards tests; androids and, if female, something like "gynodroids"? Or from another perspective; wouldn't very intelligent yet morally despicable people receive quite high scores as well? (Please remember the example of Hitler as a very good writer.)

As human learners, we are emotional creatures. If our emotive nature does indeed play an important and unique role in educating oneself and others, then certainly no exam whatsoever can actually transform our human desires themselves. Standardized and traditional exams seek some kind of objectivity; by doing so they actually pretend to be able to do away with our nature as affective subjects. Perhaps one could even go so far as to argue, as I pointed out before, that testing in a scientific setting has somehow been moved to areas in which testing is simply not the most adequate way to go around. Is the scientist with her white robe, is the doctor with her white robe, really so different from the way a teacher enters a sterilized room in which standardized testing takes place? At least scientist were robes so that we can recognize them!

But, once again, leaving aside the idealistic approach to testing, if indeed exams are here to stay and to be altered progressively, then we ought as teachers to responsibly coach our students not only to pass them, but as far as possible to excel at them. Three temporal moments ought to be considered; before, during and after. An interesting question, following Heidegger's *Being and time*, would center on what notion of time is actually used in testing. But leaving this extremely important issue aside, let us look at each.

<u>Before testing</u>: If tests are not to be considered the make-or-beak condition of education, then teachers must guide students as to what a specific test, for a particular area, might look like. This does not at all mean that the class will simply become a kind of "coaching for testing". Though

this may happen. For instance many schools in Colombia prepare their students for hours on how to pass the ICFES; our impressive Benedictian rector actually disdained this approach. It is even said that in some schools bad students are let go so that ICFES scores will be higher for a given school! Why would anybody do this? You see, standardized exams end by providing a ranking. Oddly, some think that this list is equivalent to being a good school or university. I recall registering at Carleton many years ago, for I was not sure I would enter McGill. People who saw me registering could not believe that I would do such a thing. I did not get it for I had just arrived to Canada and was free from these realistic elements. It seems as though quite a lot of self-assessment is required in education; though it would be arrogant to say that one knows how to go about it for sure. But leaving these concerns aside, I will take up two issues as regards pre-test preparation.

 First, while my students would suffer much in class because of the constant emphasis on listening and speaking, I constantly let them know that ----for the most part---- these areas would not be included in their exams (this has changed at the CCA). I was not too troubled by this discrepancy, for the most part because great concern was placed in students' actually understanding, as well, those structures and functions which would eventually appear in their exam.

 Or take the following case in the case of the Humanities. People do have a certain phobia for philosophy; I am quite sure bad philosophy teachers have had, in part, this terrible effect. In this sense most philosophers seem to many presumptuous. But some as well can provide the best, if only, examples of radical self-assessment. Some philosophies are nothing other than radical self-assessment. Having said this, while teaching introductory Political Philosophy we looked at several of Plato's dialogues. One is called the *Laches*. It deals with a very important question;. What is courage? (And people think philosophy speaks of nebulous things!) To make a long story short, before the mid-term I actually told the students that I would ask them about the conversation between Laches, Nicias and Socrates. They knew the <u>exact</u> question, for the concern in the Humanities is not to pass or fail but to transform, that is, to give another form to oneself, and perhaps, to some extent to others as well. And yet many students could not answer fully, for to be able to say what courage is; even if you read the *Laches*, that is not a simple thing to do. Socrates himself says that none of us actually know! And yet, we should, for courage is to a great extent what teachers and learners have. What I mean is this: you may prepare students

before a given exam, but in many cases such preparation does not actually result, for some reason or other, in an absolute change in the general tendency of test scores.

During testing: Having had to give and correct literally thousand of exams, I could not but agree that a good teacher gives students the best chance possible to actually write a good exam. The CCA was in this respect absolutely fair. We received a package and the very first page laid out how teachers should proceed in order for the exam to be fair for all. Such mechanical elements are for sure to be emphasized, particularly for those who are not used to taking so many exams. Not everybody's life story is that of a teacher who has been accustomed to taking exams throughout all her life. But one must really confess that some of the advise sounds truly silly. Take these words: "Advise them to concentrate as carefully as possible". Does it not truly sound quite patronizing? I am not sure, as I am not sure about many issues, but making the question need not hurt anybody. It may be that, saying it, in the end make us teachers actually feel better! These kinds of "silly" questions may even be set in place to actually lessen our guilt as we move from the "teacher friend" to what really counts, namely, the "objective judge".

  And an absolutely essential element is telling learners the time. But here I would like to tell you a story regarding test-taking strategies, stories which might emphasize the need for a certain degree of flexibility. Once at the CCA one of the learners of a very closely knit group was not able to arrive on time for her exam. At the CCA, as in many other institutions, there is for practical reasons an exact time to begin the exam. I agreed with my students that we would start later than usual to give a her a chance to arrive. But she could not make it, she told us through a cell phone. It was absolutely unbelievable to see how worried her classmates were for her. Bear in mind that there was great security regarding exams at the CCA. I told them to contact her and if she could go to my apartment I would let her present the exam right there in my dining room. They did, and breaking all examination regulations, I let her take her exam while I had dinner and watched a program on TV. Well, she did not get a good grade, but at least she passed to the next level; she truly deserved to. But I could have been fired from the CCA for this; as of yet nobody knows there about this and many other difficult decisions. In other words, during the test you must be able to respond to on-site decisions which you have to take up and not simply avoid just because you are following a certain protocol. Or in other words, have you ever been fired for defending your students? I have.

<u>After the test.</u> As I have argued above, at the CCA the final two hours of each monthly course were dedicated to providing feedback on exams. I have already mentioned how I personally dealt with the different levels of numerical achievement: high, average and low. But since each institution is quite different, then what one ought to do while working for each may require much variation. But perhaps what one should emphasize about "after the test" is that ironically what comes after the test is a certain kind of numbness on the part of our students. The learning process has been made so unpleasurable that our students need to regenerate by avoidance. All of us know how the end of a semester felt like. And the worse thing about it all, is that if you asked me how much I remember of any such exams, I must let you know that very little.

(Note: An aspect of test taking which is not even mentioned in Brown's book is the case of plagiarism. Schools, colleges and universities have strict norms on plagiarism. However, if they were to self-assess their adamant positions, perhaps they would come to realize that a legalistic attitude towards such cases simply does not get to the bottom of things. If we pressure our students to such an extent that they will do <u>anything</u> to be able to pass a given course, then something might be wrong with the structure itself. I myself can recount multiple stories of plagiarism (some very serious) and how I decided to resolve each by way of diverse ---for the most part non-legalistic--- actions. It is in the case of plagiarism that our educational system seems to show the most radical anomy to its foundations. Do you not think that Jason ----a pretty intelligent kind of kid--- would in the end be pushed to such extremes by the way the system itself works? I could provide some interesting stories in this respect, however, this assignment has again become too extended. We must move on.)

QUESTION 4

  Murray's test is quite innovative. Though I perhaps do not follow the steps themselves, quite clearly what is of central importance is the underlying objective which could generate the ability ---- as we argued above----- for self-assessment and peer-assessment. Grading others, well that might just be an "easy" affair; but grading yourself, well that is a little tougher. Who would want to fail herself? The general idea is great, to have students learn to be responsible for their own learning is what learning is all about. Learning to self-evaluate brings about greater honesty towards oneself. As part of many of my courses –where institutional restrictions made it possible---- students would give themselves their own participatory grade. Even if you never

came to class, you could give yourself a grade of excellence; for all I know you can read Aristotle on your own better than with my biased and discrete remarks on him. Moreover, learning to be corrected by others takes a lot of ability for we are all to used to correcting others rather than correcting ourselves.

But besides these two elements, two others spring to mind as being of great importance here. The first is the very initial question Murray asks, namely, the relation between the difference in the grade you would give yourself based upon what you think you <u>know</u>, and upon what the <u>effort</u> you put into it was. We have, as teachers multiple times heard how students will say —---and many honestly do so--- that they studied quite hard for the exam; they studied the most for <u>your</u> exam. And here the dilemma becomes more apparent yet; how to fail a person who has studied for hours, who has even set up long hours of group study? My own personal solution in the area of the humanities was this. Unless it was blatantly obvious that you had not done your best effort, you would be passing, or near passing. To take up a previous example. It is not easy for us so say what "courage" is. But if you have read the *Laches* and at least give me sense of what Socrates considers courage to be (though Socrates actually ends by saying that none of us know what courage really is!), then you are on the right path. (Why come up with the example of courage? Precisely because today is Remembrance Day.) However, it was absolutely certain that higher grades were not to be expected from all, for our greatest efforts sometimes actually do not produce the best of results. For example, as part of my M.A. I remember handing in my first essay. I am starting to blush. I put a lot of effort into it, but sometimes one just does not really understand what philosophical thought is all about. As I tell my students, the first penalty kick I had in my Masters, I shot way wide. Fortunately, the others I shot closer to the net! For you see, to even begin a PhD you must have at least an A- average.

In other words, passing, almost anybody can do that; getting A's, well some can do that; but reaching excellence, that is altogether another matter. Reaching excellence is something for which I believe evaluations are still to be generated. And it is no wonder that when speaking of Aristotle's ethical thought, students were reminded that in Greek *ethike arête* literally means "excellence of character", rather than simply "moral virtue" as it is usually translated. In this sense ESL testers could --maybe even should--- begin re-reading their Aristotle so that we can come up with key solutions to our testing dilemmas.

Finally, by far the most important section in Murray's test is that of the questions one finds at the end. Whatever the exercise does, if it is clear or not, if it is feasible or not, whatever these

empirical issues, the underpinning argument is absolutely crucial. Let us look at Murray's questions:

1. "After having taken this test, what kind of grade do you think you should get?:

2. Do you think this test gave a fair picture of what you know?:

3. Was it easy, fun, or what? Would you like to take tests like this?

4. Could it be improved in any way?

Asking to students to comment on exams is absolutely central if we are to move from a realistic perspectives on exam-taking to a more idealistic perspective; idealistic I must emphasize, not as in the word "illusory", but instead as understanding that human nature is capable of great things without what could in the end be seen as punishing devices. One example of fun exams. If you have ever studied philosophy, you know you have to take courses on logic. For people like me, this is quite boring. We just do not get it (or pretend not to). However, the best example of exciting logic exams was that of the son of Jorge Páramo. He, as you will recall, I defended as regards the Grammar Translation Method a few assignments ago. His son Aquiles (can you imagine Professor Páramo's love of Greek that he even named his children by using Greek names!) created some amazing exams. For instance, each and every single exam would include an extra question for those who were able to finish the exam before time. Since as you are reading you are testing me, I will test you while you read me. One of the extra points was this. Take six matches of equal length and with them create four equilateral triangles whose side is the length of one of the matches. If you get it congratulate yourself as part of you continuous self assessment! If not, I advise you to change your all too common paradigms in finding the solution and maybe you will get it. But besides, Aquiles' exams were covered with birds he himself had painted. Fortunate are we who were able to presence such pleasurable conditions; logic ceased to be a dreadful affair.

But I want to conclude this distance program by applying what we ourselves preach. Therefore let me respond this question by focusing on the distance program for TESL training itself. For I am sure you follow your own precepts and in this sense require us to provide you with feedback. In order to do so I will very briefly answer Murray's four questions, namely:

1. After having taken this test, what kind of grade do you think you should get? It's been almost two months of pleasurable yet demanding writing. I have learned, well I learned it many years

ago, that I myself do not care too much about the number or letter which you give me. Even some of my PhD professors still believe that one should care much for this. This puzzles, even troubles, me.

2. Do you think this test gave a fair picture of what you know? It depends, and this is tricky for a distance program. How to know who somebody is, from the distance; this is quite difficult. And yet we are, or better, should we not be most truly ourselves when we write? Did we not say in our previous assignment that that which differentiates the writing skill, is that it ----over all others---- generates greater self-understanding? If the test itself was fair, that I am not sure. Though Benns provided flexibility in that I had to present only 5 modules out of 12, I must confess that at times I am not absolutely convinced that I had to do it. But one does not set down the rules of the game. However, one thing is for sure, I certainly did not appreciate the arrogance of TESL Ontario. It seems as though <u>they</u>, in particular, need much self-assessment. Socioculturally speaking, Canadians are known abroad for their politeness. I hope one day I might be certified in spite of these words!

3. Was it easy, fun, or what? Distance programs are complex for in writing you will be for the most the part there with yourself. Fortunately, I seem to have fun with myself. Having a severe physical disability allows you to have fun without much movement. But if writing is to be not just readable but pleasurable -----not just well argumented, but enticing----- then I truly believe that writing must incorporate rhetorical skills such as humor and persuasions techniques. I hope you have laughed somewhat while reading these tens upon tens of pages. If not --–since we are speaking of fun things--- you can fail me!

4. Could it be improved in any way? Of course, who and what cannot. It seems to me the distance program is still quite rigid. How so? There is no flexibility in the questions to be answered, or in the order one could follow in answering them. It is true that I created some disorder in the program; but I did so at my own risk. Nowhere does it say that one <u>can</u> or even <u>should</u> proceed this way. Can you imagine if things had gone badly? I would certainly have been seen by many as quite presumptuous. And I might still be.
But moving to some concrete examples:

a. the ordering of the questions on culture is not the best. Although I pointed this out in my assignment, I felt as if little or not change could actually be made, for when discussing cultural factors the very same order was defended by the teacher of the in-classroom program without providing a real defense. That is to say, in order to know whether "culture should be taught in an SL or FL program", one need first know, at least vaguely, "what culture is". For it would be something like asking a person: Should I buy eggplant, when I do not even know what an eggplant is. Particularly with eggplant the wording itself might utterly confuse me, for it certainly is not a plant of eggs. So the order of the questions should be made more flexible.

b. The repetition of topics: As you might recall at different moments in my writing I would have to argue that such and such an issue had been treated in previous assignment quite extensively. Although for most of the issues I again provided further exemplifications (I love to learn), this may in the end hamper our ability to focus on certain issues rather than others. For example the issue of strategic competence was dealt with repeatedly. So the elimination of some questions, or parts of questions, can be thought of as a possibility.

c. Boredom. As in any testing devised, some of the exercises are quite boring for some learners. The one that tipped the scales for me was Brown's "Our Future Stock" (Teaching Reading, pages 320-330) Of course Brown had to choose one reading; however, nowhere in the program are learners moved to select their own. I did so, but again what may have turned out well, could have turned out pretty badly. So the use of your own material should be encouraged from the start.

d. alternative modes of answering: Too many discrete question seem to break a writer's writing fluidity. An example, what I wrote on the foundations of ESL teaching, was followed by some responses which for the most part required one to simply repeat the characteristics of some of the methods dealt with. Could one not do the distance program –specially in the case of experienced teachers—by writing an essay, let us say 15 pages, for each or some of the modules?
And finally,

e. In some particular cases, a more dynamic interaction between the distance program and the in-classroom program could be considered. What I mean is this; if anything I said about the Grammar Translation Method is at least partly correct, then surely this should have an effect on the in-classroom teaching. In other words, simply reducing the Grammar Translation Method to a method without a theory once again is quite discouraging for learners with a certain background in the Humanities. In this respect, I will continue to urge for the building of bridges

between ESL and Philosophy. However, at times I am not too optimistic, for the centrifugal force of one's area seems to make it difficult to pull out into regions where unknown issues seem to be explored.

Having provided this minimal and well-intended feedback, I will correspondingly conclude by assessing myself. Throughout my life I have been constantly assessed both by institutions and students. Though I have never been to keen on being assessed, the results have allowed me to grow as a teacher. But perhaps greater growth has been generated through a radical decision which involves a radical type of self-assessment in which the rigidity of most paradigms comes into question. That is, at least as far as one can question what actually supports who one is. Of course, as all of us, I am far from an adequate assessment of myself. However, instead of further focusing of this personal limitation, I will leave you by providing you with one of the best assessments which has ever been made of me. As you know, a distance program creates great distance. I do not see you, you do not see me. So that you can actually see me, look at how one of my creative students once assessed who I was:

(Some contextualization. Students were accustomed to my saying out loud, or to writing in HUGE letters, "AUCH" whenever an absolutely appalling mistake was made. Contrary to what

one would expect, they would actually enjoy it, as I enjoy looking at myself saying it in their cartoon.)

CANADIAN LANGUAGE BENCHMARKS

(If I am to take myself seriously, I have the following to say about this question. I have carefully read the introductory pages of the Canadian Language Benchmarks. If I am not to make of this assignment an incredibly longer assignment, I leave it up to you to assess whether you would rather have these 47 pages or 3 or 4 more on a topic which seems simply to ask us to repeat what is laid down in the CLB's introductory pages) (pgs. V-XIV).

QUESTION 1

  What are the Canadian Language Benchmarks 2000?

QUESTION 2

  Are the Benchmarks a type of test of curriculum?

QUESTION 3

  What are the components of the Benchmarks?

QUESTION 4

  What does a Benchmark number mean?

QUESTION 5

  What are the major features of the CLB?

QUESTION 6

  How do the CLB benefit learners?