The Challenges for Science. Education for the Twenty-First Century
Pontifical Academy of Sciences, *Scripta Varia* 104, Vatican City 2002
www.pas.va/content/dam/accademia/pdf/sv104/sv104-gueye.pdf

EVALUATION AND EDUCATION IN SCIENCE

BABACAR GUEYE

1. Introduction

Be it explicit or not evaluation always comes along with teaching (oral and written tests, exam papers, etc.) That is why it may have a central position in all well-thought of curricula.

However it is common knowledge that the evaluation part is the most difficult curriculum component to set up. That is why in Education Science it has been considered as a special field of investigation for so long.

Docimology – a subject that concerns itself with evaluation in order to disclose good and bad practices and then consolidate the former and improve the latter has gradually gained ground into evaluation research.

If many researchers have tried in the past to define the evaluation concept in a more or less successful and appropriate way, today we must agree that most of them put evaluation at the service decision making (STUFFLEBEAM 1980, De Ketele 1993).

For De Ketele, evaluation means to

– collect a set of adequately appropriate, valid and reliable information

– study the adequacy rate between this set of information and another set criteria suitable to the assigned objectives from the onset or streamlined on the way, in order to make a decision (De Ketele, 1993).

This definition corresponds to an epistemological change as if nowadays a fully documented decision making seems to be the stated objective of evaluation, the unique will to pass a value judgement from measures had seemed to overrate any other considerations for a long time.

As a mather of fact value judgement and decision making are the two stages of the same process. Any decision making stems from a value judgement on the people's actions or performance in relation to implicit or explicit objectives.

Another definition of evaluation considers that it should help determine congruence between performance and objectives that is Tyler's definition stated as follows.

"The evaluation process mainly consists in determining to which extent the education objectives principally aim at changing human beings, that is the objective is to cause desirable changes in students' behaviours, whereas evaluation is the process consisting in determining to what extent these behavarioural changes are actually occurring (Tyler, 1950)".

Therefore it could be stated that the evaluation issue is mainly rooted on the following questions

What is to be evaluated?

When and why evaluate?

How to evaluate?

Our paper which is focusing on the teaching of sciences develops within that range.

2. The Traditional functions of Evaluation

2.1. *What is to be evaluated?*

The usual answer to this question is provided by the school conception that stages the following steps in the design of any programme.

– Determine the objectives that should be aimed by the course or the programme

– Choose the learning experiences that will help reach those objectives

– Organize those learning experiences

– Determine to which extent those objectives are attained (FURST, 1964).

The point is then to evaluate the objectives, more precisely the objectives in terms of behavariours if Tyler's (1950) and FURST'S (1964) definitions are brought together.

The basic task is, therefore, to assess the objectives and categorize them. All this resulted in the already familiar taxonomies (Bloom, d'Hainaut, etc) which have, each attempted to assess, describe and categorize what the learner should be able to achieve what ever subject content may be used as support.

The first taxonomy, that of Bloom published in 1956 lists six levels (knowledge, comprehension, application, analysis, synthesis, evaluation) was used in Quebec as early as 1964, then in Belgium in 1972 for evaluating learners.

Such a method had a strong point as it rationalized systematized and evaluated an educative action which had too long been left to intuition, sensibility and common sense (De Landsheere, 1975).

However that has been a major criticism against it. The use of taxonomies entails too strong a focus of evaluation and teaching on atomised behaviours which do not take into account the initiative and desire of the learner who is thus compelled to quasi inactivity.

2.2. *When and why evaluate?*

It is customary to evaluate while learning is in progress (oral & written quizzes, progress tests, etc) or on completing a syllabus (exams, contests). Those two evaluation forms use to be kept apart by SCRIVEN (1967) who calls the first on while Training Evaluation and the second End of Training Evaluation. They are different not only in terms of the time when they are administered, but also in terms of the reason for their administration in the application of teaching programmes.

The while-training Evaluation (Progress Test) may be defined as a continuing evaluation process aiming to ensure every individual's progress in a learning strategy in a view to alter the learning situation or the rate of that progress in order to improve on remediate (if applicable) it.

The end-of training Evaluation (achievement test) which tends more and more to be referred to as Certification Test is defined as the one that leads to a binary decision for a pass or fail in relation to a learning period, for granting or denying someone a promotion, for continuing or stopping an action (De Ketele et Roegiers, 1993).

2.3. *How to evaluate*

In many countries it up to the teachers in charge of one course who design the evaluation (progress and achievement) tests and the learner is supposed to demonstrate his/her competence through a written production: it is the well-known "per and paper" test.

The oral tests are indeed administered, but writing is a given more focus with a higher coefficient.

Two main techniques are usually employed in school tests

– The so-called objective test when the learner has to choose the one correct answer from others which are not. This kind of test resulted from the early applications of taxonomies, namely in Canada in 1964 and in Belgium in 1972 as far Bloom's taxonomy is concerned.

– The composition, an answer in a written form which allows the learner to produce some more elaborate response, presented in an organised or an original free way.

In the field of sciences, the composition has developed from a traditional form in one question, one sentence or one word alone to a more structured form with several questions requiring more or less complementary answers extracted from provided documents.

That change occurred while the methodological procedure was being introduced in the teaching of sciences which allows both discovering and understanding phenomena.

3. Contributions of Didactics to Evaluation

Subject areas Didactics seems to me more comprehensive and systematic to report on the impact of evaluation on the teaching of Sciences, which is the reason why I have chosen it as a scope for this discussion.

If Didactics was originally, indeed, a new approach to educational issues, it has to be noticed that it has presently expanded beyond the school field. Now it deals with all the communication settings, be they formal, non formal or unformal.

Coming back to the school setting which the object of our concern, let us consider that the Didactics of a given subject area both looks at classroom proceedings and at what happens in the learner's mind.

In short, it has to do with the way messages are encoded and transmitted, in priority, but most principally with "how learners learn" and how they interact with the learning contents and the teacher's strategies.

Didactics positions itself at the crossroads of the three following domains.

– The subject area domain (programme – contents – objectives)

– The psychological domain

– The pedagogical domain

It calls upon each of these domains if need be, to give a definition and meaning to the school tasks, depending on the obstacles facing the conception and acquisition of knowledge and skills. Today, didactics has identified two short comings in the teaching of Sciences.

First school has generally restricted evaluation to the sole field of learning.

Second, the quality of the evaluation battery in use is so poor that it has negative consequences on the whole curriculum

Be it in the form of a progress or an achievement test, evaluation in the teaching of sciences is only geared towards learner acquisitions. However, as far as the teaching field is concerned there are other contests which require the collection of reliable and valid data before decision making.

– First, it is now common knowledge that actual teaching is preceded by the teacher's awareness of his learner's ideas. Which will help him take into account libely problem areas when planning lessons.

– Second, the teaching tools used as supportive materials learning bear information and values that are worth disclosing for optimal efficient use.

– Third the different evaluation results also include useful information on learner behaviour and how it operates.

As mentioned above, evaluation in the teaching of Sciences is today confined to the results of acquisitions. Given this situation, the Didactics of Sciences, through its research results has revealed thanks to a more systematic vision of the teaching act, other fields, other instances when evaluation may play an important role in the quality of learning of teaching aids and reinforcement that may be granted as shown in the grid below.

EVALUATING MEDIATION

MEDIATOR	WHAT TO EVALUATE
1. Knows the targets	Characteristics, questions, interest, conceptions, thought procedures
2. Has objectives	Knowledge Know-how Attitudes
3. Has communication resources	Teaching aid Films Books Posters Teaching modules
4. Has constraints	Space Time Funding, etc…

3.1. *Diagnostic evaluation*

If the teacher pays some attention to "errors" made by learners on such or such a concept or scientific reasoning, he realises that some of those are made again and again on a regular basis.

It was logical for didactics to look closer at errors. This is how researchers discovered that prior to a course on a given topic learners had a number of ideas on it, their own explanations of some phenomena and interpretation of the environment surrounding them. Those prime ideas, those rough elements in the learner's brain that most often opposed to settled scientific knowledge are referred to as conceptions or representations. The sciences didacticians started evaluating those conceptions. If the learners' conceptions issue has been raised since the works of Piaget, its systematic study started in France only with the works of Pr. Giordan (1975, 1977, 1978) who came to realise like other researchers that learners tended to forget most of the scientific knowledge acquired at school or in other terms the "pedagogic yield" that is the amount of knowledge acquired in relations to the time spent at school is very weak, even non existant, at times. Presently over (w) three quarters of studies published in sciences didactics deal with conceptions they fall into three categories.

– *Descriptive researches*

They assess learners' conceptions and draw up questionnaires, etc, kinds of catalogue. Unfortunately, they wank, I believe, as the most numerous.

– *Explanatory researches*

They go beyond mere categorisation of conceptions, they aim at identifying the mechanisms that generate them and how they operate.

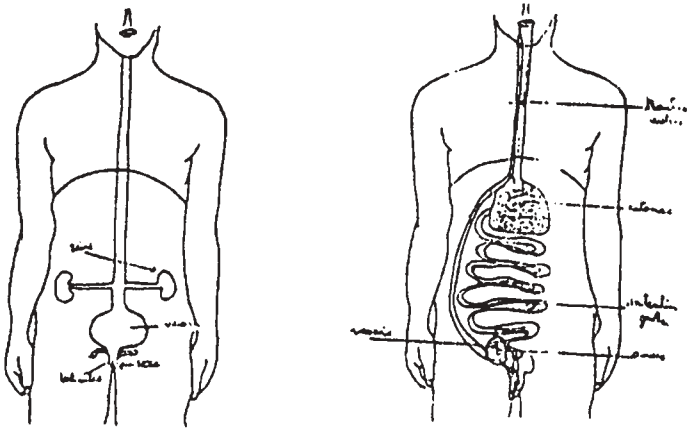
– *Applied researches*

They are few however they seems very important to me as they try to install teaching strategies usable in class and taking into account learners' conceptions.

All these researches are credited to put the lear back at the beginning and end of the education act and have made obvious what follows.

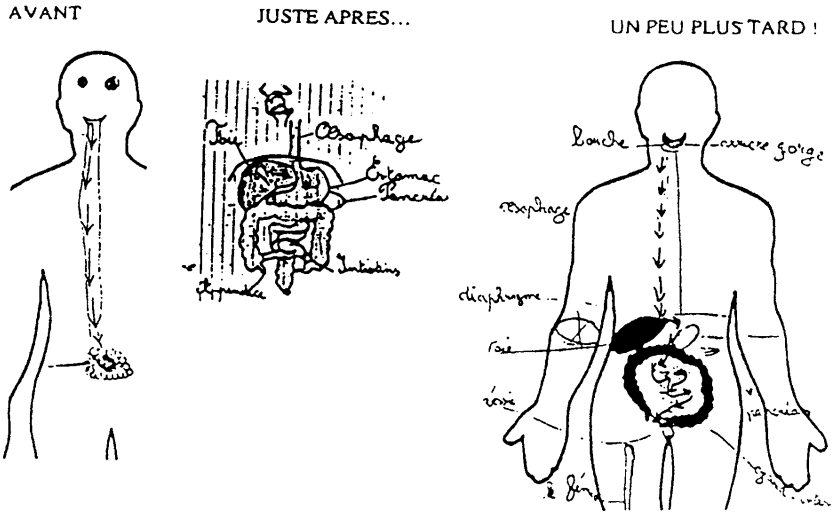
– *Before any teaching* learners have conceptions, ideas or reference framework allowing them to capture the different messages.

Here are a few examples about the digestive apparatus (Giordan, 1988).



1 2

Document 1. Two different conceptions of the digestive apparatus. In Case 1 There is a confusion with the excretion apparatus and the continuing part of the oesophagus leading to the bladder. In Case 2. Two canals (one of them would represent the trachea) One for solids – This shows that the learner is always active he always functions with prime ideas at back of his mind. Conceptions are extremely difficult to eradicate. If the teacher does not take them into account what is taught will only transit, as shown in the example below still about the digestive apparatus (Giordan, 1988).



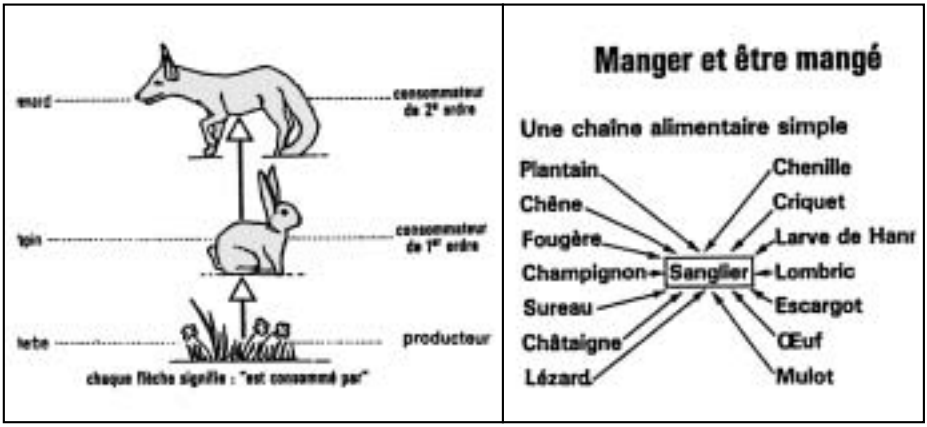
Document 2. Evolution of a conception of the digestive apparatus after the lesson.

3.2. *Evaluation of Teaching aids*

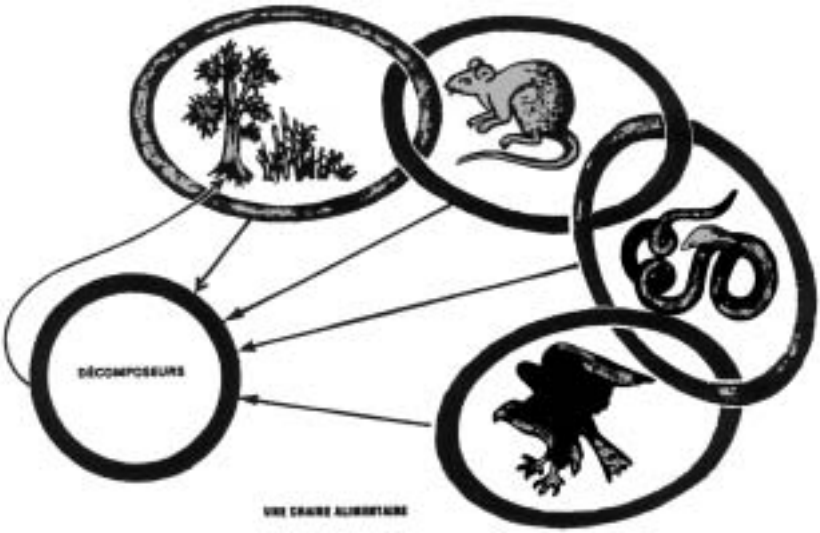
In the teaching of experimental sciences, the teacher often resorts to aids of icaried kinds in order to create situations that favour actives learning. This how he moves from exposure to demonstration at to problem solving at times. In such situations they help the teacher reach move easily the targeted objectives allowing learner to build up knowledge by themselves. They also permit to engage into activities which would not be possible, other wise.

For a number of years now, teaching research has been stressing the fact that teaching materials (books, films) carry information and even values that deserve evaluating before use not only avoid embarring situations but to assess their efficiency. Here are a few examples taken from school text books.

EXAMPLE 1.

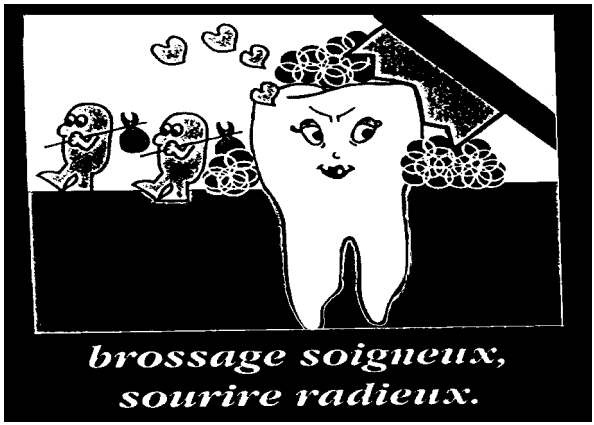


Document 3. The food chain in ecology. If the drawing on the left may be receivable for its simplicity, that on the right is hardly acceptable. Moreover, judging from terminology familiar to learners, this has nothing to do with a “chain” – “Chain” refers to a sequence of rings as shown below.



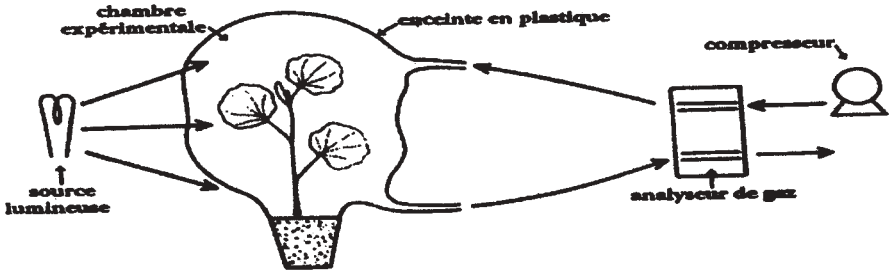
Document 4. Another representation of the food chain concept. But here, the arrow does not mean "is consumed by". This is a semantic and conventional puzzle which needs to be evaluated and straightened up for a better understanding of the food chain concept.

EXAMPLE 2. Information about tooth hygiene.

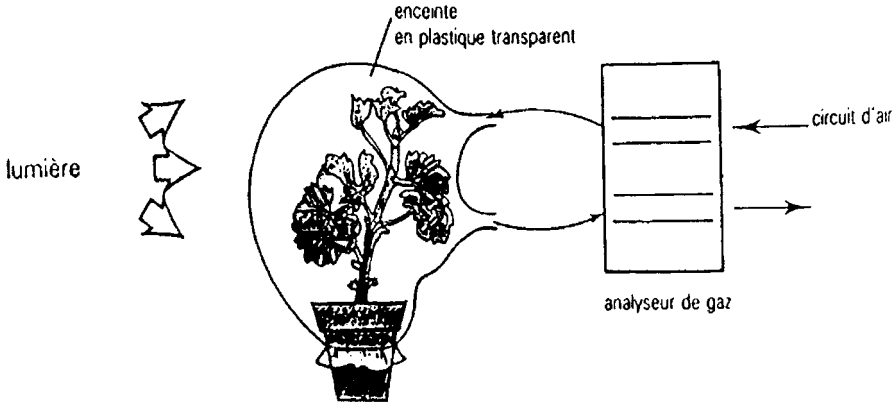


Document 5. Children-geared information.
– The adjective "delightful" is not part of the 8-10 year old children's lexicon.
– The tooth brush is spotted by less than one child out of two.
– Bacteria are not identified by 75% of children.
– Worse most children even cannot identify the tooth.

EXAMPLE 3. Measuring gas photosyntetic exchanges.



Document 6. Graph of principle of the gaz photosynthetic exchange measure.



Document 7. Another graph of the principle of the gas photosynthetic exchange measure.
First of all, it can be said that despite a few symbolic differences the two graphs taken from different text books seem to represent the same experiment.
However a closer does not take into account of all sorts of gazes which might due to the numerous organisms of the soil, be emitted and disturb the composition of air within the cover. Little cause in the teaching materials great effects on the reasoning and rigorous, procedure which are characteristics of experimental action.

3.3. *Evaluation of evaluation tools*

If one goes beyond the good wishes expressed by official texts on the evaluation of learning in science in search of habits settled and the very practice of the different tests designers, one discovers that in the whole, the pathology of evaluation in sciences goes beyond the long-critised subjectivity of judgements passed by examiners. It is founded on its chronical lack of validity relative to the stated contents and objectives, validity being defined by the extent to which tests evaluate what they are supposed to evaluate.

In many countries, teaching is centred around acquired knowledge. The starting point is university knowledge already built up, then, a list of themes followed by a list of teaching contents which rank high on the programmes of the educational system.

In the case of achievement Evaluation in Sciences, when a categorisation by themes of topics suggested over a fairly long periods is carried out and compared to the prescribed programme, it can be noticed, that in most cases in practices, all the programme, it can be notices, that in most cases in practice, all the programme chapters are not given equal treatment.

This allows to find out that there is a sort of an implicit value scale according to which certain chapters are more frequent. Others occur from time to time while one category never appears. A times a clear cut discrimination between the different notions and concepts can be notices within the same chapter.

A systematic study of objectives from the questions usually asked in sciences tests shows that contrary to the state and documented desire to move science teaching towards experimental procedure and scientific attitudes

Questions that require the use of acquired knowledge are the most frequent, in spite of the misleading appearances of the different instructions used by designers in setting questions (Analyse, Interpret, Deduce).

In fact learners are asked to show what follows:

– specific knowledge

– mastery of one (or more) problem solving (strategy[ies]), type(s)

– (exceptionally) skills to analyse a situation under study (Johsua, 1983).

Let us take the example of the evaluation of the learners' skill to form hypothese. It is noticed that the existence of familiar laws whose demonstration has necessitated hypothese known to the learners after they have been taught in class does not leave much room to questions evaluating their skills to form their own hypothese.

Indeed, when learners are faced with such questions, they look back into the set of explanations they already known for those that can fit, as they are never asked to exclude any hypothesis, but to choose the best.

Ultimately, the whole issue is only related to constructed automatisms solely for the monitoring of knowledge in the frame word of a closed systems whose importance is nothing else than success at the exam (Johsua, 1983).

As for the affective objectives, they have not all been evaluated (it must be agreed that they are difficult to evaluate in the framework of a "pen and paper" test).

The consequences are that, after a certain time, the examination turns round typical questions which are the object of swotting such as the exam tasks of analysis, interpretation or even explanation will require declaratives or procedural knowledge transfer.

This is mainly explained by the fact that all the given exercises are put forward in the frameworks of a pen and paper test, which does not allow learners to face a real problem, that is a more or less new situation whose answer must be built up not retrieved automatically.

Moreover, in the countries where, for sometime, people have cherished the idea of making learners acquire, about the themes on the programmes, certain basic concepts and procedures from one or two examples, then to make them show during the evaluation what they have learnt though a transfer to a context not dealt with in class, have waken up from that dream.

It is noticed rather an insidious knowledge inflation phenomenon in the chapters focused on at the exam, as every time a new example came up at the tests, it was systematically swotted on the following year in class on teachers' more initiative finally, overwhelmed by a mass of observations, subjects or definitions, most learners have no other solutions than learn as many of them as possible by heart. The sequence observation-memorization.

Monitoring and oversight becomes frequent practice instead of the opposite: observation-use of concepts-interpretation-relation with wider concepts-problem solving (Novak, 1970).

4. Conclusion

At the end of this analysis I think to have widely proved that it is quite possible and useful, and even urgent and compulsory to widen the evaluation field in the teaching of sciences to the differentes questions that underlie the evaluation issue and the need to make decisions only when fully

informed widely claim for such expansion. The adoption and the practice of such a systemic vision in scientific teaching will without any doubt, allow to improve it qualitatively. This improvement is dependent up on the quality of it's the learning evaluation tools.

In the teaching of sciences, as it seems presently, achievement testing is more debatable than progress testing.

Indeed the pseudo-democratic centralism used to administer in a more transparent an fairer way the evaluation of objects resources does not permit, during the achievement test, to muster all the didactic available to the scientific disciplines in their every day teaching in every school institution, owing to the number of candidates and the limited time allowed to the exam.

To solve this problem of evaluation in the teaching of sciences it is desirable to remember that whatever the philosophy and the content of a syllabus, its actual efficiency is largely determined, after a number of years, by the exam format for which the teacher has to train his learners (Guinier, 1980).

I believe that, for a project aiming to encompass all the ground aspects of the evaluation in the teaching of sciences to be likely to get off the it is necessary to tackle the problem through teachers' initial and continuing training.

Indeed teachers' straight forward training in theoretical, but mainly practical and critical training is the challenge to meet for evaluation to occupy its rightful position in the teaching of sciences.

BIBLIOGRAPHY

Bloom *et al*. (1969). *Taxinomie des objectifs pédagogiques*. Lavallée trad; Montréal: Edition nouvelle.

De Ketele, J.M., Roegiers, X. (1993). *Méthodologie du recueil d'information*. Bruxelles, De Boeck, p. 74.

De Landsheere, V. et G. (1975). *Définir les objectifs de l'éducation*. Liège, Thone, p. 15.

Giordan, A. (1975). "Analyse des attitudes scientifiques et des représentations des élèves". *Bulletin INRP*, n° 11.

Giordan, A. (1977). "Libération et conceptualisation". *Revue française de pédagogie*, n° 40.

Giordan, A. (1978). *Une pédagogie pour les sciences expérimentales*. Paris, Centurion.

Giordan, A., De Vecchi, G. (1987). *Les origines du savoir, des conceptions des apprenants aux concepts scientifiques*. Neuchatel, Delachaux et Niestlé.

Giordan, A., De Vecchi, G. (1989). *L'enseignement scientifique, comment faire pour que ça marche?* Nice, Z'Editions

Giordan *et al.* (1993). *Evaluer pour innover.* Nice Z'Edition

Guinier A. (1980). "A propos du contrôle des connaissances". *Bulletin de l'Union des physiciens*, n° 627, pp. 29-72.

Johsua S. (1983). "Contrôle des connaissances en fin de second cycle (bac-calaureat) et nouveaux programmes de physique". *Revue française de pédagogie* n° 64, pp. 55-73.

Novak, J.D. (1970). *The improvement of biology teaching.* New York, Bobbs Merill and Company.

Scallon, G. (1988). *L'évaluation formative des apprentissages*. Québec, Presses de l'Université LAVAL.

Scriven, M.S. (1967). *The methodology of evaluation* (AERA Monograph series on curriculum evaluation, book I). Chicago, Rand Mc Nally and Co.

Stufflebeam, D. *et al.* (1980). *L'évaluation en éducation et le prise de décision*. Ottawa, Edition NHP, p. 48.

Tyler, R.W. (1950). *Basic principales of curriculum and instruction, Syllabus for Education 360*. Chicago, University of Chicago Press.