# Metacognitive indices for realism in self-assesment

*Dieudonné Leclercq(1) & Marianne Poumay (2)**

*2006*

**University of Liège**
(1) Professor. Head of the Service de Technologie de l'Education (D. **Leclercq@ulg.ac.be**)
(2) Professor. Head of the LabSET-IFRES (**M.Poumay@ulg.ac.be**)
http://www.labset.net

## Introduction

Analysing the titles published in the research literature, Tochon (1991) observed that the thema of metacognition had its peak in 1988 and then a drop. A similar peak-then-drop phenomenon had been observed with a related concept in the '70s : confidence marking. This "wave" phenomena are not surprising when we consider the opacity and inoperability of many definitions, the absence of consensus on the main concepts, and, last but not least, the lack of valid and efficient instruments and methods having a high consequential validity, i.e. an important impact on learning.

This is a pity when we consider the issues related to these concepts. For instance, from their meta-analysis on the researches dealing with the factors that influence learning Wang, Haertel & Walberg (1990) conclude that the most important one is metacognition. Nevertheless, we can still make the same observation as Bereiter & Scardamalia (1989, 380) : "there is little support for students in developing knowledge about knowledge". This paper will present an operational definition of metacognition leading to a taxonomy of the metacognitive activities, and then focus on one of the possible approaches : Degrees of Certainty.

**Metacognition** should be considered as a reflection of a person on one's cognition in a **multidimensional space**, with at least five dimensions : the situations (learning or testing), the temporality (before, during or after the situation), the objects of reflection (the productions or the processes that lead to it), the level of consciousness (for the learner) and the level of observability (for the trainer), the operations (judgement, analysis and regulation). Several definitions are compatible with this paradigm; they must be judged at their usefulness (to produce new knowledge) and practicability (to be used in real situation). That is why we are ready to change the provisory version of our current **operational definition** of metacognition :
« *Observable judgements, analysis and /or regulations effectuated by a learner on his/her own performances (learning processes or products), in situations of PRE, PER or POST performance (mainly testing or learning).* » (Leclercq & Poumay, 2003).
The Taxonomy can be represented as follows :

| | PRE | PER | POST | |
|---|---|---|---|---|
| Judgement | | **Degrees of Certainty** | | |
| Analysis | | | | |
| Regulation | | | | |

The current presentation will be dedicated on one of those possible modalities of metacognition, i.e. Degrees of Certainty, located in one cell of the model.

The **Degrees of Certainty** principle **(previously called Confidence marking)** consists in asking to a student to add a Degree of Certainty (DC) to each answer to a test and to score according to the student's realism in addition to the correctness of his/her answers. The reasons for adopting such a process will be explained, as well as the ways to collect these new data (the instructions given to the learners or the testees) and of exploiting them, including new ways of marking, i.e. scoring tests.

The rationale underpinning the whole approach has been coined by De Finetti (1965) in the expression "partial knowledge" and his claim : "*Partial knowledge exists. To detect it is necessary and feasible*" (p. 109).

Leclercq (1983, 1993, 2003) has regularly presented synthesis of the experimental study of Degrees of Certainty (DC) or confidence marking started fourty years ago (Van Naerssen, 1962 ; Shuford & al., 1966 ; Rippey, 1968 & 1970 ; Hambleton & al., 1970 ; Jacobs, 1971 ; Pitz, 1974 ). This research movement has suffered of epistemological opacities and has consequently produced theoretical confusion and unexploitable data.

During the last 30 years we have tried (Leclercq, 1983, 1993, 2003 ; Leclercq & Poumay, 2003) to operate epistemological clarifications (of what is knowledge, for instance), to conceive appropriate indices of personal realism, to accumulate experimental evidences (on human subtleness and realism in the use of the probability scale for instance). The presentation and discussion of those principles will constitute the core of the present paper.

**Principle 1 : The purpose of DCs**

The ultimate **purpose** of the study of the use of Degrees of Certainty (DC) in school settings is to help students improve their learning processes and their uses of knowledge. It will be shown that Degrees of Certainty (DC) are linked with data gathering behaviour, with change of responses, with quality of performance. Whereas it consists in judgements during the (testing) situation, in order to benefit to learning, it should be followed by analysis and regulation after the (testing) situation, for instance in e-mail supported "metacognitive dialog" with students based on metacognitive indices, more specifically realism indices (see below) derived from their use of confidence degrees. Extracts of such dialogs will be provided afterward.

**Principle 2 : Human limitations in DCs**

Like other human capacities, the use of Degrees of Certainty (DC) is characterised by general **properties and limitations** shared by the majority of humans and by specific traits, some belonging to the persons and others belonging to the situations. This implies that observations (data that can be collected) result from the interactions of these parameters. Among others, we have studied (Leclercq, 1983, 1993) the limits of the human sensitivity (or granularity or subtleness) to estimate their chances of producing the correct answer on a probability scale. We discovered that (untrained) adults could hardly distinguish in a reliable way more than 7 (plus or minus two) Degrees of Certainty (DC) on the scale ranging from 0% to 100%. This lead us to use instructions asking to the students to choose among only 6 Degrees of Certainty (DC) : 0%, 20%, 40%, 60%, 80%, 100%. This scale is symmetrical and does not include 50%.

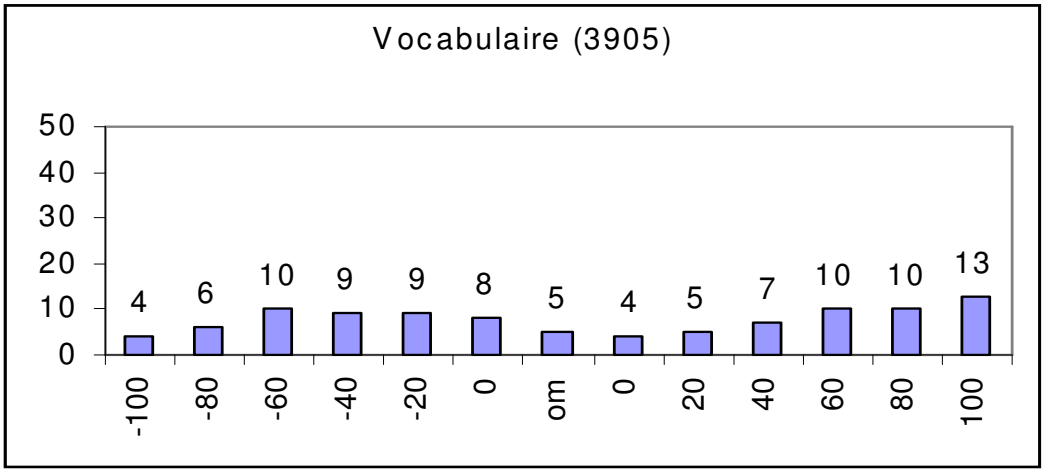
This characteristic will be helpful when we will exploit the couples "Answers + DC" and derive indices of Realism.

**Principle 3 : The Degree of Certainty (DC) is a part of the definition of knowledge**

Degrees of Certainty (DC) and derived metacognitive indices are **based on a theory of knowledge and performance** that assumes that the degree to which a person believes his/her answer is correct constitutes the very definition of knowledge…correct or incorrect (misconceptions for instance). In DeFinetti's terms, « *It is Only subjective probability that can give an objective meaning to every response and scoring method* » (1965, 111).

The degree of quality of knowledge can be represented (graphically and mathematically) as a "spectral" continuum of quality of responses to a test. This continuum is illustrated here after by one of the **spectral distributions** obtained during the MOHICAN study (Leclercq, 2003) where about 4000 students entering 8 universities[1] have taken 10 tests (Vocabulary, Syntax, Text comprehension, graphic comprehension, math, biology, physics, chemistry, History-economics-actuality and Arts). Questions were 7 alternatives multiple choice (5 classical + the "None" solution, + the "All" solutions). In addition to their responses, students had to give one of the six confidence degrees described here over.

The following figure represents the spectral distribution of 175725 answers (45 answers and 45 DCs given by 3905 students) to the Vocabulary test (45 questions). The worse performance (on the extreme left side) is having given an incorrect answer with the maximal Degree of Certainty (DC), code -100%. It has happened for 4% of the anwers. The best performance (on the extreme right of the spectrum) is a correct answer given with the maximal Degree of Certainty (DC), coded as +100% or 100%. It happened for 13% of the answers.




Hunt (1993) suggested to distinguish between three types of knowledge situations in which a person can be in relation to a piece of content : "misinformed, uninformed, informed". We would place them respectively on the left, middle and right side of the continuum. More specifically, misinformed knowledge (or misconception or misleading knowledge) correspond to the incorrect answers given with a Degree of Certainty (DC) higher than 50% (60%, 80% and 100%). This kind of knowledge could also be called "dangerous knowledge". Informed knowledge correspond to the correct answers given with a Degree of Certainty (DC) higher

[1] Of the French speaking part of Belgium.

than 50% (i.e. 60%, 80%, 100%). This kind of knowledge could also be called "usable knowledge". In between those two extreme blocks, the rest can be considered as representative of "unusable knowledge" since the person has not enough confidence in it to apply it in a reliable way.
It is alarming that in this vocabulary tests students entering the universities of this country display only 33% of usable knowledge, 47% of unusable one and 20% of dangerous one.
The quality spectra differ from student to student, from test to test, from boys to girls, from sections to sections. All those data are exposed in the MOHICAN book (Leclercq, 2003).

The same kind of data has been also collected in domains related to aircraft piloting and maintenance (Leclercq, 1982), to health and especially to urgence medicine and diabetic patients in Liège (Leclercq et al., 2003), in Padova (Brutomesso et al., 2003) and in Paris (Reach et al. 2005). Following, George Bernard Shaw's principle

"Beware of false knowledge : it is more dangerous than ignorance".

**Principle 4 : Confidence degrees enable subtle measures of gains due to learning**

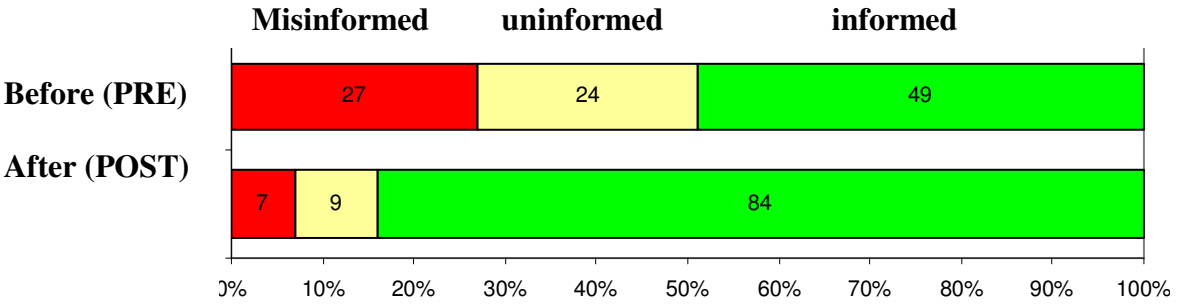
Jans (1999) has computed systematically changes in performance at a test before and after a learning session (where students could explorate a hypermedia course). Here is a typical result (for student 50) on a 100 item test :

| | Pre | Post |
|---|---|---|
| usable K | 58 % | 89 % |
| unusable K | 35 % | 6 % |
| dangerous K | 7 % | 5 % |
| | 100 % of items | 100 % of items |

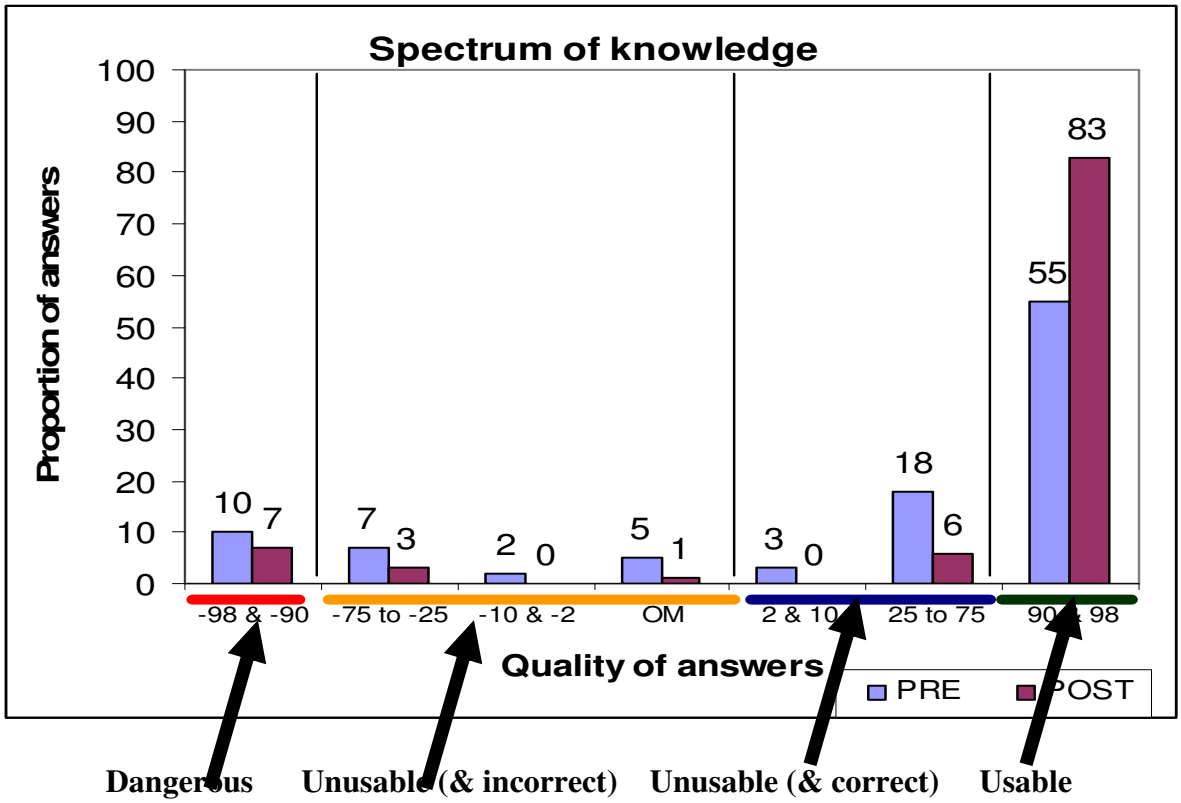
Learning has diminished the rate of dangerous knowledge (from 7 % to 5 %) but, essentially, has increased the rate of usable knowledge (from 58 % to 89 %).
The strategic importance of this kind of gains have for long been encapsulated by Mark Twain:

"It is not what we do not know that harms us.
It is what we believe to be true and isn't."

In a similar experiment, one of our students, Lucas (2001) tested 300 students aged 8-12 before and after a video related to first help behaviours in case of accidents. Here are her observations, presented in a graphical way :



In an other context, patient education, diabetic patients have been tested about their knowledge on diabetis and its treatment in different settings (Leclercq, Rinaldi & Ernould, 2003) before and after training sessions. Here are the overall pre and post-tests results from Brutomesso et al. (2003) for 38 patients having answered 39 items :




In Lucas' and Brutomesso et al. 's data (with an outdated way of collecting Degrees of Certainty), the remaining 7% of misinformed (dangerous) knowledge at the post-test should be addressed in priority since "the most useful piece of learning for the uses of life is to unlearn what is untrue".

Antisthenes (445-365 B.C.)

**Principle 5. Convenient indices of realism are necessary and possible**

Metacognition is a complex process and its expressions (such as Degrees of Certainty degrees) are composed by intricated variables of different natures : **cognitive** ones, **affective** ones and **conative** ones that can be conceptually distinguishable and should be measurable separately in order to give diagnostic feedbacks to the learners.

We have described elsewhere (Leclercq, 1982) the series of indices of realism that have been developed to address these issues. The most known are
The "Realism By Calibration index" (RBC)
The "Mean Error of Centration index" (MEC)
The "Internal Coherence index" (ICI)

MEC is easy to compute, but a good value (0 error of Centration) may result from the compensation of overestimations by underestimations. RBC and ICI are difficult to compute ; they are ambiguous and cannot been analysed clearly. Since these indices provide hardly

helpful tools for educational use, we have developed (Leclercq & Poumay, 2003, 181-190) indices of realism that are simpler to compute and easier to interpret, keeping a high sensitivity to changes in behaviour :
The **Confidence** index is the average confidence degree given with correct answers.
The **Imprudence** index is the average confidence degree given with incorrect answers
The **Nuance** index is the difference between those two indices.
We refer to these three values as our CIN indices of realism. We also use CPN since we mean that a person who has alow Imprudence index is Prudent.

Here are the average values of those CIN indices for the 10 MOHICAN tests (Leclercq, 2003, 47) :

| **Total** | Vocab | Syntax | Text Compr | Graphs Comp | Math | Phys | Chim | Biol | Arts | Hist-éco |
|---|---|---|---|---|---|---|---|---|---|---|
| **Confidence** | **60** | **73** | **68** | **62** | **78** | **69** | **66** | **63** | **63** | **59** |
| **Imprudence** | **40** | **56 *** | **48** | **40** | **51*** | **41** | **38** | **43** | **30** | **46** |
| **Nuance** | **20** | **17** | **20** | **22** | **27** | **28** | **28** | **20** | **33** | **13** |

It can be seen that students have been the most Confident (78%) for the test in math, but that they have also been highly imprudent (51%) in this same test. It is in artistic knowledge that students have demonstrated the highest degree of Nuance (33%) since they also demonstrated the lowest imprudence in this same test. Briefly, it can be said that "in artistic knowledge, students can distinguish clearly between when they know and when they do not know".

**Principle 6. Scoring systems should reinforce realism**

For each content, these CIN indices should be given **threshold values**. For general culture contents (such as history, geography, arts, etc.), we consider that confidence should be (strictly) higher than 50%, imprudence (strictly) lower than 50 % and nuance (strictly) higher than 20%.
In other contents such as medical urgency, aircraft piloting and aircraft mechanics, we fixed the confidence threshold at a 100% value, and for diabetic patients, at the 90% value.

In the assessment of our students in various universities (Liège, Paris, Aosta), and in domains such as educational psychology and technology, we score the students as follows.
-The "classical" score is obtained by giving 1 point per correct answer and by withdrawing 0,25 point per incorrect answer, regardless of the number of distractors2 or whether it is an open ended question or a MCQ.
-The total is transformed on a 20 points scale.
-The Confidence index is computed as the average value of the Degrees of Certainty given with correct answers.
-The Imprudence index is computed as the average value of the Degrees of Certainty given with incorrect answers.
-Only if the student has given a confidence degree for each of his answers, we add
0,5 point if Confidence > 50%
1 point if Confidence > 60%

[2] Usually, correction for guessing in Multiple Choice Questions depend on the number of alternatives and distractors.

1,5 point if Confidence > 70%

,

0,5 point if Imprudence < 50%

1 point if Imprudence < 45%

1,5 point if Imprudence < 40%

NB : Whereas those three last "bonus" are based on the Imprudence index, it is Prudence that is rewarded (i.e. having a low value at the Imprudence index).

This gives to the students a clear idea of the importance we attribute to metacognition (that never penalises !!!!). Interviewed, they declare to be satisfied with this scoring method and understand the epistemological and social values underpinning his approach.

The following table presents scores obtained in the written exam (with multiple choice questions) of a course in 2003 on Higher Education given by D. Leclercq to university freshman in psychology. In the first row ("Total"), are presented the average value (over 300 students) of the classical score (9,3/20) and of the final score (combining correctness and realism), i.e. 11,3/20. .

| | Realism | | Metacognitive Bonus | | |
|---|---|---|---|---|---|
| | Imprudence | Confidence | Imprudence | Confiance | Total |
| 587 | 40 | 58 | 1 | 0,5 | 1,5 |
| 597 | 53 | 64 | 0 | 1 | 1 |
| 1509 | 40 | 65 | 1 | 1 | 2 |
| 1557 | 43 | 73 | 1 | 1,5 | 2,5 |
| 1864 | 47 | 54 | 0,5 | 0,5 | 1 |
| 10503 | 55 | 75 | 0 | 1,5 | 1,5 |
| 10886 | 8 | 49 | 1,5 | 0 | 1,5 |
| 10998 | 47 | 64 | 0,5 | 1 | 1,5 |
| 11467 | 46 | 72 | 0,5 | 1,5 | 2 |
| 11490 | 57 | 75 | 0 | 1,5 | 1,5 |
| 11638 | 20 | 60 | 1,5 | 0,5 | 2 |
| 11660 | 32 | 78 | 1,5 | 1,5 | 3 |

This table shows that, for this exam, in average, students have won 2 points on the 20 points scale, where the passing score is 12. One student (the last one) won 3 points, the maximum.
Not obtaining the Confidence score is rather exceptional, but it happens (see student 10886 who is very prudent : only 8% as Imprudence index).

On about 300 students, average Confidence was 68%. Average Prudence was 46%. Average Nuance was 22%.

This scoring system is based on a **polychotomous principle** : the Confidence bonus are obtained or not (the same with Prudence and the students are concerned "to have it" (the extra point) and are disappointed when they have not obtained it ; they check why and try to come up with a new strategy to improve the quality (realism) of their self assessment in the future.
We consider that this discontinuous system (here 6 bonus values) has a good consequential validity and attracts students' attention on the thresholds of qualities of realism.

During years we have used scoring formulas that integrate cognition and metacognition, with 12 Tariffs (see table hereafter), 6 TCs (Tariffs in case of Correct answer)and 6 TIs (Tariffs in case of Incorrect answer).

| Degree of Certainty | 0 | 20 | 40 | 60 | 80 | 100 |
|---|---|---|---|---|---|---|
| Correct | 9 | 12 | 14 | 16 | 18 | 20 |
| Incorrect | 2,5 | 2 | 1 | -1 | -6 | -20 |

They are based on decision theory (Leclercq, 1982), where the expected score at each question (ESQ) of a student is

ESQ = $(p \,.\, TC_{DC}) + (q \,.\, TI_{DC})$, whre

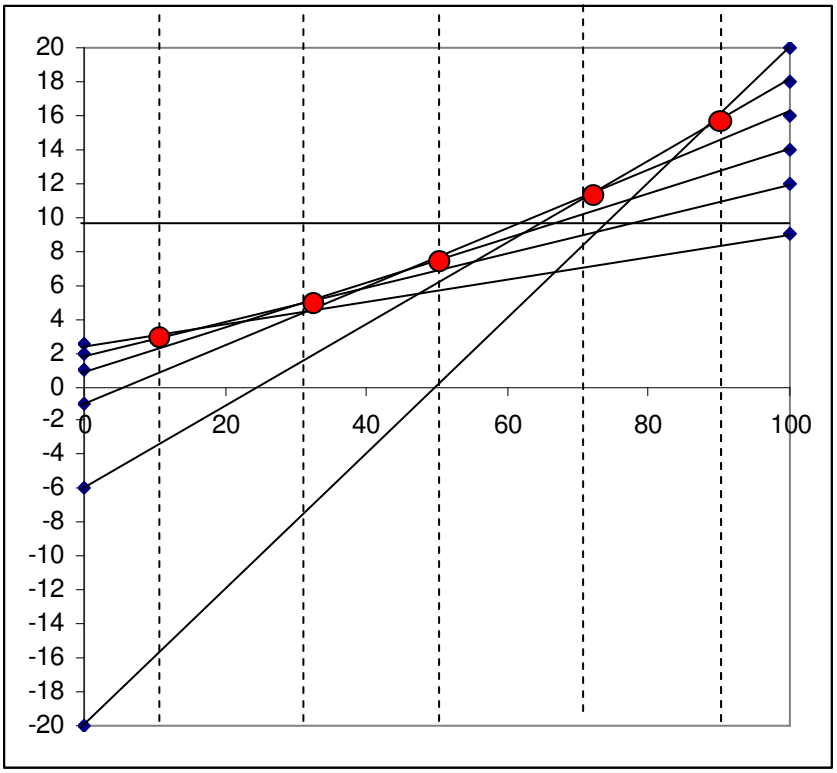
p is the probability of giving the correct answer

q is the probability of not giving the correct answer (omissions + incorrect) ; actually, q = 1-p

$TC_{DC}$ is the TC (tariff) for the DC given

$TI_{DC}$ is the TI (tariff) for the DC given

For each couple (TC and TI) of values fixed for each of the six degrees of Certainty (0, 20, 40, 60, 80, 100); a resulting ESQ value can be computed for all the values of p and q. this constitutes an oblique line. The six oblique lines have been plotted on the graph hereafter. One can observe that each line overtops all the others (is maximal) at the vertical of the probability region (horizontal axis) where the given Degree of Certainty is recommended by the instructions. For instance, the line associated to DC = 60% is maximal in the zone ranging from 50% to 70%, those limits being indicated by bullets. This scale of 12 tariffs has been computed to make sure that the student can maximise his expected score only by saying the truth or by expressing his best estimate without bias for each answer.



Unfortunately, we observed that this contributed to opacify the system and to encourage students to play Montecarlo games such as "give always the 60% degree of confidence, regardless of your answer", i.e. playing the game of maximising the total score without considering one's certainty at each question. Those strategies take into account the scores, not the probability of each answer being correct. The students were playing a quite different game than the one we invited them to play.
Our split approach, rewarding separately correctness and realism in self assessment encourages the students to assess their certainty at each answer without bias. It has an other important advantage : to provide pieces of information easily computed : the Confidence and the Imprudence indices, the values of which are independent from the correctness.

Improving the scoring system by making it more and more valid (theoretical, predictive and consequential validity), informative and friendly has been a continuous process during those last 35 years (we started it in 1971). Students reactions helped to discard a lot of inappropriate ones. We would welcome any contribution to this movement.

There is a price to pay for simple systems : **the extreme situations** have to receive appropriate solutions. For instance when a student has given only correct answers, Imprudence cannot be computed neither rewarded and the classical score (before any bonus) is already 20/20. Therefore, we first credit this student with a extra bonus of one point for this perfect performance (at the objective level) and give him an extra bonus of 1,5 point, than compute his Confidence Bonus. With such a system, a student who has the maximum number of correct answers has the possibility to keep his advantage in confidenced scores in comparison to a student who made one error. This subtlety is important in case of normative selection procedures (ranking the students) where only the x (say 200 for instance) students with the highest scores are selected.
Nuance is not rewarded per se since it is already rewarded via the Confidence and Prudence Bonus. Rewarding Nuance instead of the two other indices would be unfair when the Confidence an Imprudence indices are both higher than 50% or both lower than 50%.

We end this section with a good news : realism can be trained and improves with experience (Leclercq , 1993, 129 ; Leclercq & al., 2006). We are engaged in **metacognitive dialogs** with students, by mail, starting with the metacognitive indices obtained from formative evaluations. In this dialog, we start by inviting the student to comment the values of their indices, with the questions "Why ?" ( i.e; stepping into the analysis process) and "What would you change ?" (stepping in the regulation process described in our definition of metacognition). This research is very promising when we consider the students' capacity to analyse, to diagnose their mental processes and to make regulation decisions.

**Principle 7 : Degrees of Certainty and Metacognitive indices should help understand learners' behavior**

In a famous poem, T. S. Eliot's says :

"Where is information we lost in **data** ?
Where is knowledge we lost in **information** ?
Where is **wisdom** we lost in **knowledge** ?"

Eliot's terms can be paraphrased as follows. Before acting, people often suspend judgement and look for data, either to understand, or to answer, or to act or to solve problems. Only some of the encountered data constitute information, i.e. "what reduces uncertainty"[3]. Only a part of

[3] According to Shannon'and Weaver's (1949) definition .

information will be maintained in long term memory and integrated into knowledge. And we all know that knowledge should be used under the control of wisdom, and that only a part of knowledge helps in improving wisdom.

In addition to their formative value, Degrees of Certainty (DC) are particularly appropriated for researching practitioners. Thos locution is coined to mimic Schon's (1988) expression "reflexive practitioner". Researching practitioners are teachers who prioritize their educational function, but that manage opportunities to make research on the real life data.
We already mentioned the domains where they are of practical use (aircraft pilots and mechanics, urgency, patient education). They are also used with university students whose professionality involves the capacity to learn and to reflect, especially about one's cognition. This of course includes any university student. Nevertheless these metacognitive capacities should be trained as soon as in primary school.

At a more fundamental level of research on mental processes , DCs are convenient. Their explicative and predictive characteristics (combined with Learning Strategy) in those processes have been demonstrated in various experimental settings by our team (among others) :

The revision of subjective probabilities about an event after the reception of an informative message, and comparison of human estimations with Bayes' theorem. (Edwards, 1967 ; Leclercq, 1983, 270)

The decision to search for information and its effects: Leclercq & Boskin (1990, 19)

The observation of subtle changes in mental structures or opinions of mind. (Leclercq & al., 1999 ; Leclercq et al, 2002)

Gender differences in over and underestimation (Leclercq, 2003, 72-91).

The self estimation capacity in young students (11 year old) (Toby, 1995).

The relation between the spectral quality of performance at pretest and the Relative gain after learning. (Lardenoye, 1995, Jans 2000).

**The evolution of the mastery in a content as well as in general competencies.**

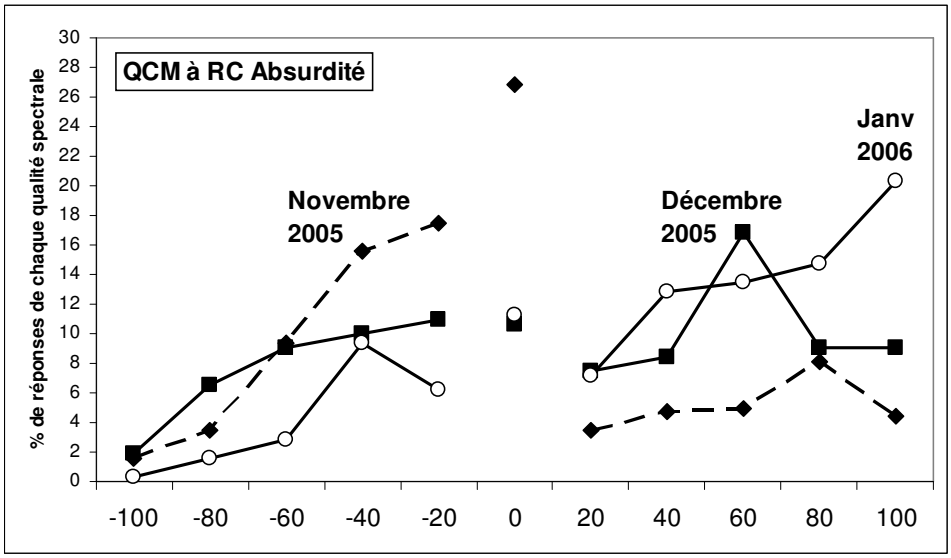
We will comment this last point by citing a recent experiment (Leclercq et al., 2006)we made by presenting to the students tests made with Multiple Choice Questions with General Implicit Solutions, namely

6-"None of the above solutions"
7-"All of the above solutions"
8-"Missing data so that more than one (but not all) solution can be correct (but not simultaneously)"
9-"Absurdity in the stem making the question (and, therefore answering) non sense".

These numbers have been chosen to fit with optical reader system. Our MCQs never present more than 5 solutions (to which the 4 GIS must be added).
These solutions are named "General" because they are the same for any question of the tes. They are called "Implicit" because they are announced at the start of the test, but not repeated in each question, so that they train (and test) cognitive vigilance.
Our freshmen are not at all prepared to these General Implicit Solutions. The first time they use them, they have bad results. Training improves the performance.

In order to train students, 2 simulations of full size tests have been presented to them in November 2005 and in December 2005, the Exam being in January 2006.

The graph hereafter reports only about the MCQ for which the correct answer was the GIS "Absurdity'. It has been drawn comparing the results of 80 students who have attended the 3 tests. It can be seen that in November (dotted lines), the rate of correct wrong responses was low (right hemispectrum) and that they improved in December and January, not only in detection, but (especially in January in their confidence in their capacity to detect.




**Conclusion**

Everybody acknowledges the importance of metacognitive skills and mathetic4 competencies for nowadays learners. Self assessment is only an aspect of them and Degrees of Certainty are only one way among others to address the issue. We hope that this article has demonstrated that some restrictions that are legitimately associated with this technique are carefully taken into account, that there are valid and reliable ways to use Degrees of Certainty, and that it has demonstrated that this technique offers the potential for new and fecund approaches to old problems. We have decided not to enter the debate of the definition of competency, since place was lacking here and we wanted to focus on technical aspects. Nevertheless, we are confident that this approach can bring its special light in the old debate so well stated by an Arab proverb[5] :

"He who knows and knows that the knows is wise - follow him
He who knows not and knows not that he knows not is a fool - shun him
He who knows not and knows that he knows not is a child - teach him
He who knows and knows not that he knows is asleep - awaken him".

## References

Bereiter, C. & Scardamalia, M. (1989). Intentional Learning as a goal of instruction. ? In Resnick L.B. (Ed.), Knowing, Learning and Instruction. Essays in honor of Robert Glaser. (pp. 361-392). New Jersey – Hillsdale : Lawrence Erlbaum Associates, Publishers.

[4] Word coined by Gilbert (1962) from the greek word μανθανω (I learn), to designate « in relation to learning ».

[5] From Leclercq & Bruno, 1993, Introduction.

Bruno, J. (1993). Using Testing to Provide Feedback to Support Instruction : A Reexamination of the Role of Assessment Organizations, in D. Leclercq & J. Bruno (Eds). Item Banking : Interactive Testing and Self-Assessment. NATO ASI Series, Berlin : Springer Verlag, pp. 190-209.

Brutomesso, D., Gagnayre, R., Leclercq, D., Crazzola, D., Busata, E., d'Ivernois, JF., Casiglia, E, Tiengo, A. & Baritussio, A. (2003) The use of degrees of certainty to evaluate knowledge. Patient Education and Counseling, 51, 29-37.

De Finetti, B. (1965). Methods for discriminating levels of partial knowledge concerning a test item, British Journal of Mathematical and Statistical Psychology. 18, 87-123.

Hambleton, R., Roberts, D & Traub, R. (1970). A comparison of the reliability and validity of two methods for assessing partial knowledge on a multiple choice test. Journal of Educational measurement. 7, 75-82.

Hunt, D. (1993). Human self assessment : theory and application to learning and testing, in D. Leclercq & J. Bruno (Eds). Item banking : self-assessment and interactive testing. NATO ARW, F112, Berlin : Springer Verlag, 177-189.

Jacobs, S. (1971), Correlates of unwarranted confidence in responses to objective test items, Journal of Educational Measurement, 8, 1, pp. 15-20.

Jans, V., Leclercq, D.(1997). Metacognitive realism : a cognitive style or a learning strategy ?, *Educational Psychology*, Special Issue of Learning Styles and Strategies, 17, n°1 & 2, march 1997, 101-110.

Jans, V. & Leclercq D. (1998). Mesurer l'effet de l'apprentissage à l'aide de l'analyse spectrale des performances. In Depover & Noel (Eds). L'évaluation des compétences et des processus cognitifs : Bruxelles : De Boeck

Leclercq, D. (1982). Confidence marking, its use in testing, 106. in Postlethwaite & Choppin, Evaluation in Education, vol. 6, 161-287, Oxford : Pergamon Press.

Leclercq, D. (1985), Computer-managed Testing (943-944) & Confidence Marking (967-969), in T. Husen & N. Posthlethwaite, The International Encyclopedia of Education, Oxford : Pergamon.

Leclercq, D. (1990), Computer-managed Testing (453-454) & Confidence Marking (343-345), in H. Walberg & G. Haertel, The International Encyclopedia of Educational Evaluation, Oxford : Pergamon.

Leclercq, D. & Boskin, A. (1990), Note taking behavior studied with the help of hypermedia, in Estes, Heene & Leclercq (Eds), Proceedings of the 7th International Conference on Technology and Education, Bruxelles. Edimburgh : CEP Consultants, 2, 16-19.

Leclercq , D., Simon, F., Marotte, P., Verschueren, A. et Lacaille, C. (2002). Former des étudiants de première candidature universitaire à des compétences transversales : Lesquelles et comment ? Deuxième Congrès des chercheurs en Education de la CFWB, Louvain-La-Neuve.

Leclercq, D. et Bruno, J. (Eds) (1993), Item banking : self-assessment and interactive testing, NATO ARW, F112, Berlin : Springer Verlag .

Leclercq, D. et al. (1993). Validity, reliability and acuity of self-assessment in educational testing, in D. Leclercq & J. Bruno (Eds). Item banking : self-assessment and interactive testing, NATO ARW, F112, Berlin : Springer Verlag, 114-131.

Leclercq D., Rinaldi A.M. et Ernould C. (2003). Un questionnaire spectral pour l'évaluation des connaissances chez le patient diabétique, in Gagnayre et al. (Eds), L'évaluation de l'Education Thérapeutique du Patient, Paris : IPCEM.

Leclercq D. & Poumay, M. (2003). Analyses édumétriques et indices métacognitifs appliqués aux questions des 10 check-up MOHICAN. In D. Leclercq (2003), Diagnostic cognitive et métacognitif au seuil de l'université. Editions de l'université de Liège.

Leclercq, D., Herremans, T., Englebert, L. , Feyens, C., Lourtie, E., Malengrez, D. , Séverine Delcomminette, Castaigne, J.L. & Poumay, M.(2006). Résultats expérimentaux d'une méthode (CAMEL) d'entraînement et d'évaluation des étudiants de première année universitaire. 23° Colloque de l'AIPU, Monastir.

Lucas, C. (2001). Evaluation des connaissances d'enfants en matière d'urgence avant et après le visionnement d'une vidéo d'information. Mémoire de licence en Sciences de la santé publique. Université de Liège.

Pitz, G. (1974), Subjective probability distribution for imperfectly known quantities, in L. W. Gregg (Ed.), Knowledge and Cognition, New York : Wiley.29-41.

Reach, G. , Zerrouki A., Leclercq , D. d'Ivernois, JF. (2005). Adjusting insulin doses : from knowledge to decision. Patient Education and Counseling., 56, 98-103.

Rippey, R. (1970) A Comparison of Five Different Scoring Functions for Confidence Tests**.** *Journal of Educational Measurement*, Vol. 7, No. 3, 165-170.

Shannon, C.E. & Weaver, W. (1949), The Mathematical Theory of Communication, University of Illinois Press.

Shuford, E., Albert, A. & Massengill, N. (1966). Admissible probability measurement procedures. *Psychometrika*, 31,125-145.

Tochon, F. V. (1991). Les critères d'expertise dans la recherche sur les enseignants. *Mesure et Evaluation en Education.* Vol. 14(2), pp. 57-81.

Van Naerssen, R.F., A scale for the measurement of subjective probability, Acta Psychologica, 1965, 20, 2, 159-166.

Wang, M. C., Haertel, G. D. & Walberg, H. J. (1990). What influences learning ? A content analysis of review litterature. Journal of Educational Research, 84(1), 30-44.