Objective Assessment of Subjectivity :
Degrees of certainty and Partial Knowledge

communication presented at the EARLI SIG on « Metacognition »
at Amsterdam on june 30 –july 2 2004

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

both from the University of Liège
(1)Service de Technologie de l'Education (D. Leclercq@ulg.ac.be)
(2) Head of the LabSET (M.Poumay@ulg.ac.be)

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, what Kirk (1997) calls "practical significance".

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 the shadowed box 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).

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 <u>judgements</u> during the (testing) situation, in order to benefit to learning, it should be followed by <u>analysis</u> and <u>regulation</u> after the (testing) situation, for instance in e-mail supported "metacognitive dialog" with students (Leclercq & Wislez, 2004) 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 <u>human sensitivity</u> (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

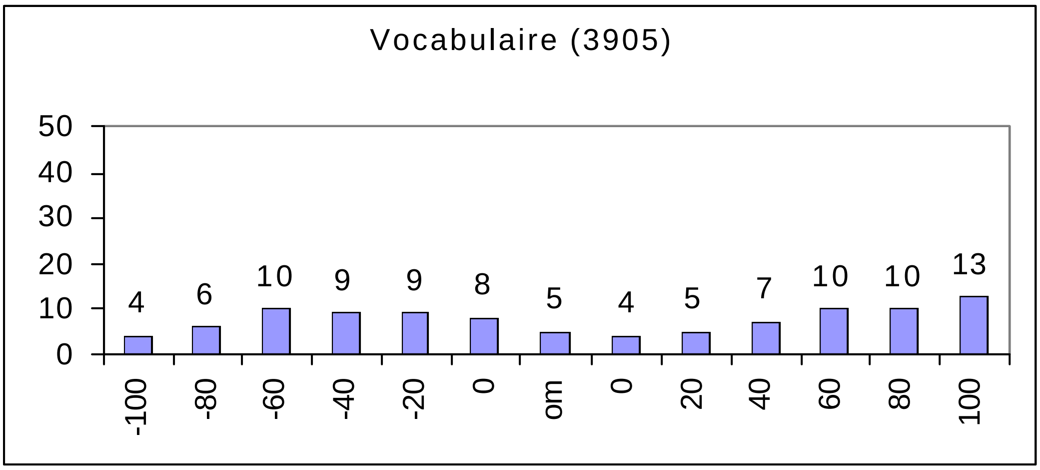
(DC) : 0%, 20%, 40%, 60%, 80%, 100%. This scale is symmetrical and does not include 50%. This characteristic will be helpful when we will exploit the couples "Answers + DC" and derive indices of Realism.

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 heare 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%. 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, 1983), to health and especially to urgence medicine and diabetic patients in Liège (Leclercq et al., 2003) and in Padova (Brutomesso et al., 2003).

Following, George Bernard Shaw' principle
"Beware of false knowledge : it is more dangerous than ignorance",

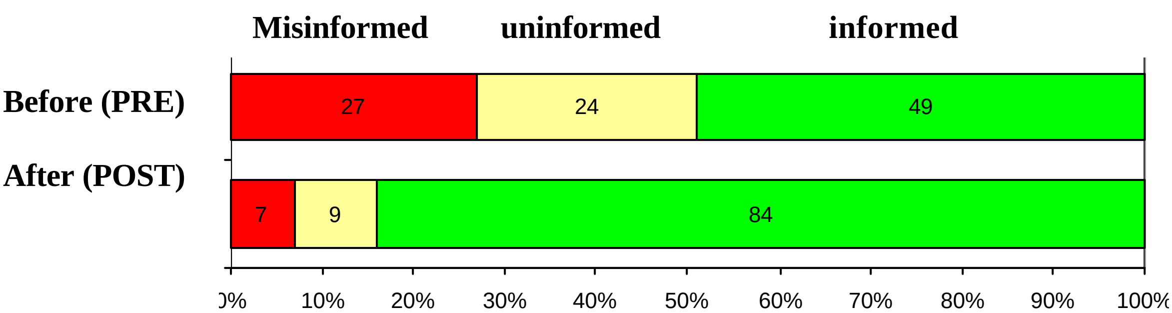
Principle 4 : Confidence degrees enable subtle measures of gains due to learning

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 :

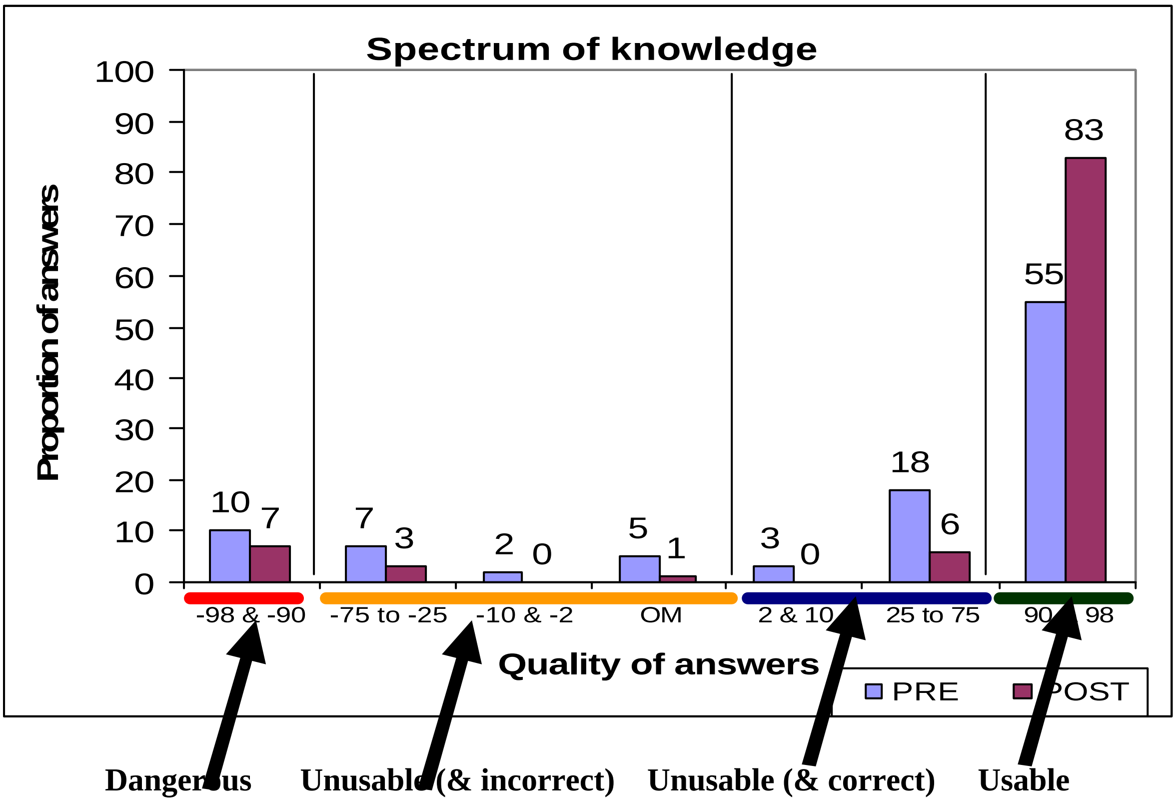
	Pre	Post
usable K	58 %	89 %
unusable K	35 %	6 %
dangerous K	7 %	5 %
	100 % of items	100 % of items

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 :

In a similar experiment, 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, 1993) 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's data, 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 confidence 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, 1983) the series of indices of realism that have been developed to address these issues.
Grounded in the works of Brier (1950), Adams & Adams (1961), Oskamp (1962), Murphy (1972, 73, 74), Lichtenstein *et al.* (1997) presented the "state of the art" of ONE approach. Its main components are the following ones :

The "Realism By Calibration index" (RBC) (the formulas and norms of which can be found in Leclercq (1993, 127-130) is the direct application of the historical trend evoqued hereover. RBC expresses the proximity of the Observed Success Rates (OSRs) to the announced ones , i.e. the Confidence Degrees, or Predicted Success Rates (PSRs) in a complex way (degree by

degree) , often done by computer since it is long to compute it by hand. Moreover, its meaning is difficult to interpret since it provides only the "Mean Confidence error" (20% for instance) but in an absolute value, not indicating whether the error is made by underestimation or by overestimation. For these reasons (not practical and ambiguous) it can hardly be (and has hardly been) the basis of metacognitive dialog and reflexion.
The "Mean Error of Centration index" (MEC) is the difference between the average Confidence Degree and the Objective Success Rate at the test (The percentage of correct answers). When MEC is negative, it means underestimation , when it is positive, it indicates overestimation and when its value is 0, it means perfect centration. This index is easier to compute than RBC, but a same value (0 or perfection) can result from patterns of behaviours largely different, from being realistic with each of the 6 confidence degree to the compensation of huge overestimations by huge underestimations. The ambiguity of this index can be compared to the claim of a person to be "on time, in average" since he is always either 30 min late or 30 min before the time of the date. Nevertheless, the predictivity of academic success in the MOHICAN study (Leclercq, 2003, 149-150) is higher than the RBC's one. In the same way, we discovered (Leclercq, 2003, 97) a rather interesting stability (a 0,43 average correlation index) of MEC across different tests in the same students.

The "Internal Coherence index" is the correlation between confidence degrees and success rates and indicates how far the student is coherent with him/her self. The problem with such an index is that a person can be highly coherent, but largely unrealistic.

Since those indices provide hardly helpful tools for educational use, we have developed (Leclercq & Poumay, 2003) 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 :

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 surgery, aircraft piloting and aircraft mechanics, we fixed the confidence threshold at a 200% value. For diabetic patients, at the 90% value.

In the assessment of our students in various universities (Liège, Paris, Aosta), 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 distractors[2] or whether it is an open ended question or a MCQ.
-The total is transformed on a 20 points scale.
-To this “classical” score, only if the student has given a confidence degree for each of his answers, and if his confidence degrees are not all the same, we add
1 point for Confidence, 1 for Prudence and 1 for Nuance if the respective thresholds are reached (the Nuance bonus can be obtained only if the two preceding bonus have been won).

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 values underpinning his approach.

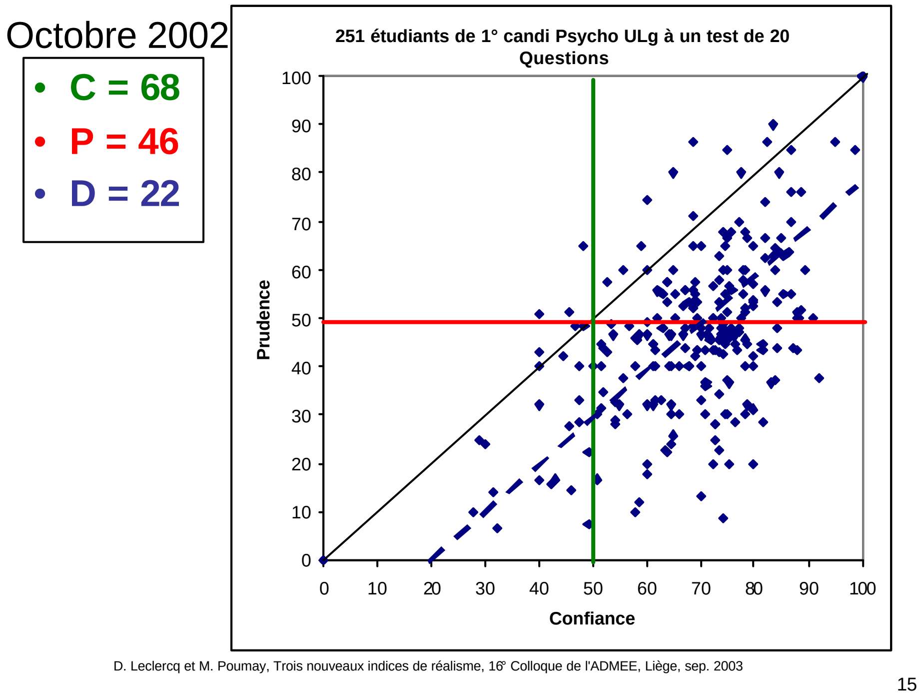
The following table presents values obtained in the written exam (with multiple choice questions) of a course in 2003 on Higher Education givan by D. Leclercq to university fresshman in psychology. In the first row (“Total”), are presented the average value (over 300 students) of the classical score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

Student code	Confidence	ImPrudence	Nuance	Conf. Bonus	Prudence Bonus	Nuance Bonus	Classical score	Final score
Total							**9,3**	**11,3**
A	**58**	**40**	**18**	**1**	**1**	**0**	**7,9**	**9,9**
B	64	53	10	1	0	0	15,3	16,3
C	65	40	25	1	1	1	10,5	13,5
D	73	43	30	1	1	1	8,9	11,9
E	54	47	7	1	1	0	15,3	17,3
F	75	55	20	1	0	0	13,7	14,7
G	49	8	42	0	1	0	7,4	8,4
H	64	47	18	1	1	0	4,7	6,7
I	72	46	26	1	1	1	8,9	11,9
J	75	57	19	1	0	0	10,5	11,5

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.
Three students (C,D& I) out of 10 obtained the 3 metacognition points, the final score of permitting them to pass. Not obtaining the Confidence score is rather exceptional, but it happens (see student G who is very prudent : only 8% on the imprudence index).

The overall situation can be presented in a single graph :

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



Average Confidence was 68%. The students plotted on the right of the vertical line received the Confidence bonus. Average Prudence was 46%. The students plotted under the horizontal line received the Prudence bonus. Students who are in this left down corner of the area and under the dotted oblique line received in addition the Nuance Bonus (Average Nuance was 22%).

This scoring system is based on a dichotomous principle : the Confidence point is obtained or not (the same with Prudence an Nuance). Other principles can be considered, more polychotomous. For instance, give an extra 0,5 point for a Confidence index between 50% and 60%, an extra 1 point when it is comprised between 60% and 70%, 1,5 point between 70% and 80%, 2 points when higher than 80%. The same reasoning can be applied with Prudence and Nuance. A continuous principle can also been considered, consisting in the use of a formula of the type +((Conf Index-50)/10) applied only if Conf is (strictly) higher than 50. In the case of this formula, the bonus will range between 0 to 5 points. If it is judged as excessively generous, it can be moderated by a denominator. For instance, it could be +((Conf Index-50)/20) where the bonus ranges only from0 to 2,5. The same continuous principles could be applied to Prudence and Nuance. We did not engage into these ways of scoring since the justified concern of being fair to each student is likely to contradict consequential validity. As we experimented, with a dichotomous system, the students are concerned "to have it" (the extra point) and is disappointed when he has not obtained it, checks with and tries to come up with a new strategy to correct his estimation answers in the future. Having used during years polychotomous systems (with 12 values), 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", since it is what pays the most, with the least risk. 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. That is why we stick to a dichotomous approach, but we go on considering improvements of the scoring system by making it more subtle, provided it does not impair consequential validity (i.e. the effects it has on students behaviours).

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, Prudence and Nuance cannot be applied 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 point if his average confidence degree is higher than 50% and a +2 bonus if his average confidence degree is higher than 60%. 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 importance in case of normative selection procedures where only the x (say 200 for instance) students with the highest scores are selected.

As can be seen, we are still looking for a good balance between equity and efficacy. We end this section with a good news : realism can be trained and improves with experience (Leclercq , 1993, 129). We are engaged in **metacognitive dialogs** with students, by mail, starting with the CIN indices or their formative evaluation. In this dialog, we start by inviting the student to comment the values of their CPN indices, with the question “Why ?”. This research is very promising when we consider the students’capacity to analyse, to diagnostics of their ental processes and to make regulation decisions.

Principle 7 Metacognitive indices can be useful in psychometrics

In the MOHICAN study (Leclercq, 2003,97), the stepwise multiple regression analysis revealed that variables including the expression of subjective probabilities were more frequently incorporated into the multiple regression equations than variables that did not include any subjective component to predict academic success of students. There were variations from faculty to faculty. In particular, the Error of Centration score and the Total Mean Spectral Score (this one, MS, having been computed only for research purposes) that consists in giving a score of +40 for a correct answer with Confidence degree 40% and -40 for an incorrect answer with a 40% Confidence degree.
The Multiple regression predictors of academic success (in first year) for the faculties of Letters (N = 154 students) are :
Syntax **MS**, History **MS**, Compreh. of graphics RS, with a Multiple R = 0,38.
The two first predictive variable incorporate the confidence degrees.

For veterinarian students, the predictive variables of academic success are
Math **MS**, Chemistry **MS**, Compreh. of Text RS and Compr. Of Text **MEC** with a RM = 0,55.
Three of the four predictive variables incorporate the confidence degrees.

From the same data bank (the MOHICAN study), Gilles (2002) showed that Spectral Bisrerial correlation indices improves item analysis, permitting a more subtle detection of flaws in alternatives of multiple choice questions.

Indices for item analysis incorporate now systematically the Confidence index for the correct answer and the Imprudence indices for each distractor. The following example (from Lebrun, Lega & Leclercq., 2003) shows the interest in entering in such details since in math (and physics and Syntax), student tend, in average to be largely imprudent (that is noted by a star when the Imprudence index is higher than 50%).

Math : Manque de Prudence

Pierre a actuellement 6 ans de plus que Paul. Il y a 10 ans, Pierre avait un âge double de celui de Paul. En supposant que Y représente l'âge de Pierre et X l'âge de Paul. Choisissez les équations représentatives du problème.

55,30%	%	r.bis	CM	
1. Y = X + 6 et 2Y = X	3,9	-0,19	56,9*	
2. Y = X - 6 et 2Y - 10 = X - 10	4,5	-0,21	59,6*	
3. Y = X + 6 et Y = 2X - 10	8	-0,15	68,1*	
4. Y = X + 6 et (Y - 10) = 2(X - 10)	**70,5**	**0,43**	**78,4**	Confiance
5. Y = X - 6 et (X - 10) = 2(Y - 10)	5,9	-0,15	64,4*	
6. aucune	4	-0,12	49,5	
7. toutes	0	XXXXX	XXXXX	

D. Leclercq et M. Poumay, Trois nouveaux indices de réalisme, 16e Colloque de l'ADMEE, Liège, sep. 2003

23

Principle 8. **Confidence degrees can be useful in edumetrics**

The expression "edumetrics" has been coined by Carver (1974) to designate the aspect of each test that can provide information on intra personal changes. Unfortunately, the number and power of tools specific to edumetrics are not yet numerous. Mac Guigan contributed with the concept of Relative gain and De Ketele (1978) with his MUCER indices. We hope we have contributed to this effort by the present paper reminding years of efforts to clarify concepts, principles, tools and practices. The main contribution should be considered in the increase of validity of measures in terms of extracting new information from the test data : not only on cognition but, in addition, on metacognition, with, hopefully, a great impact (consequential validity) on learning.

From an edumetric point of view, we would like to end with the painting of a research landscape, where some landmarks already exist (we will mention some) but where a large area has to be further explorated. This area is the explicative or predictive power of confidence degrees on learning behaviour, especially if combined with metacognitive reflections on one's learning preferences, strategies and capacities. Other papers should be necessary to develop this theme. Let us just initiate it by a poem and some related experimental results.

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 **wiseness** 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 information will be maintained in long term memory and integrated into knowledge. And we

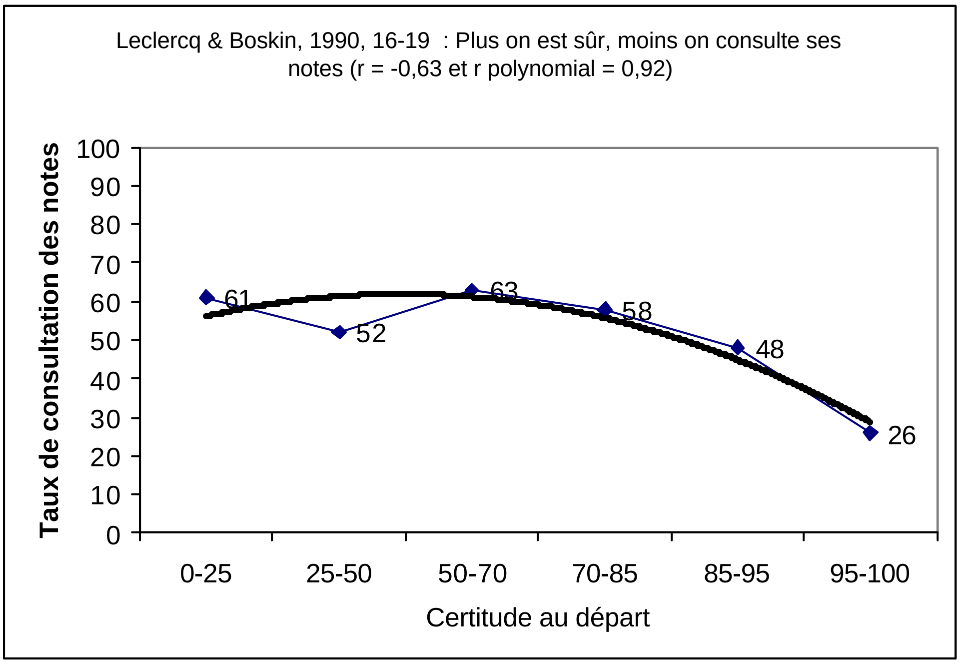
[3] According to Shannon'and Weaver's definition .

all know that knowledge should be used under the control of wiseness, and only a part of knowledge helps in improving wiseness.

Explicative and predictive characteristics of Confidence Marking (CM) and of Learning Strategy (LS) in those processes have been demonstrated in various experimental settings such as the following ones.

(CM) decision to search for information : Leclercq & Boskin (1990, 19) observed that the more the student doubts (on a pretest), the more likely he seeks for information if it is possible.

When confidence degree was…	0%	20%	40%	60%	80%	100%
…the rate of resources consultation was	61%	52%	63%	58%	48%	26%

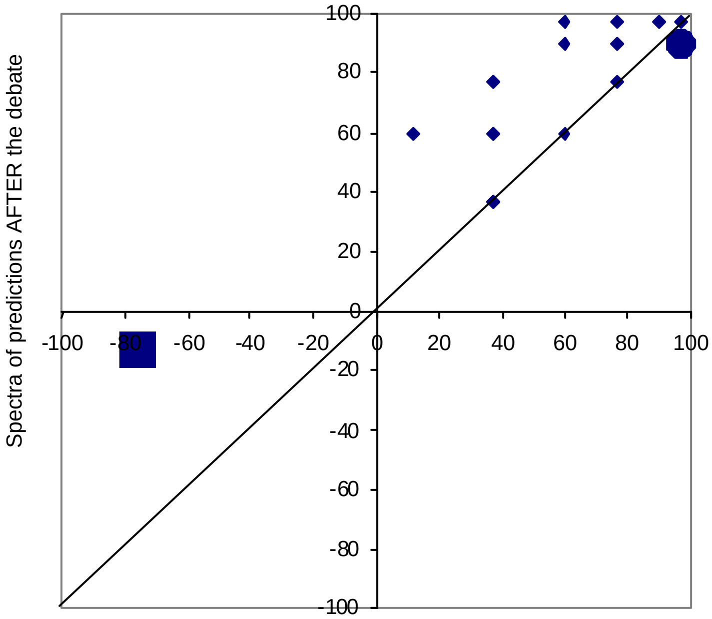


These results support Descartes' views (1636) that "doubt is the incentive of knowledge" (pp. 126-127 of the 1952 edition). The consulting behaviour is explained by subjective reasons, not by the "objective" state of knowledge.

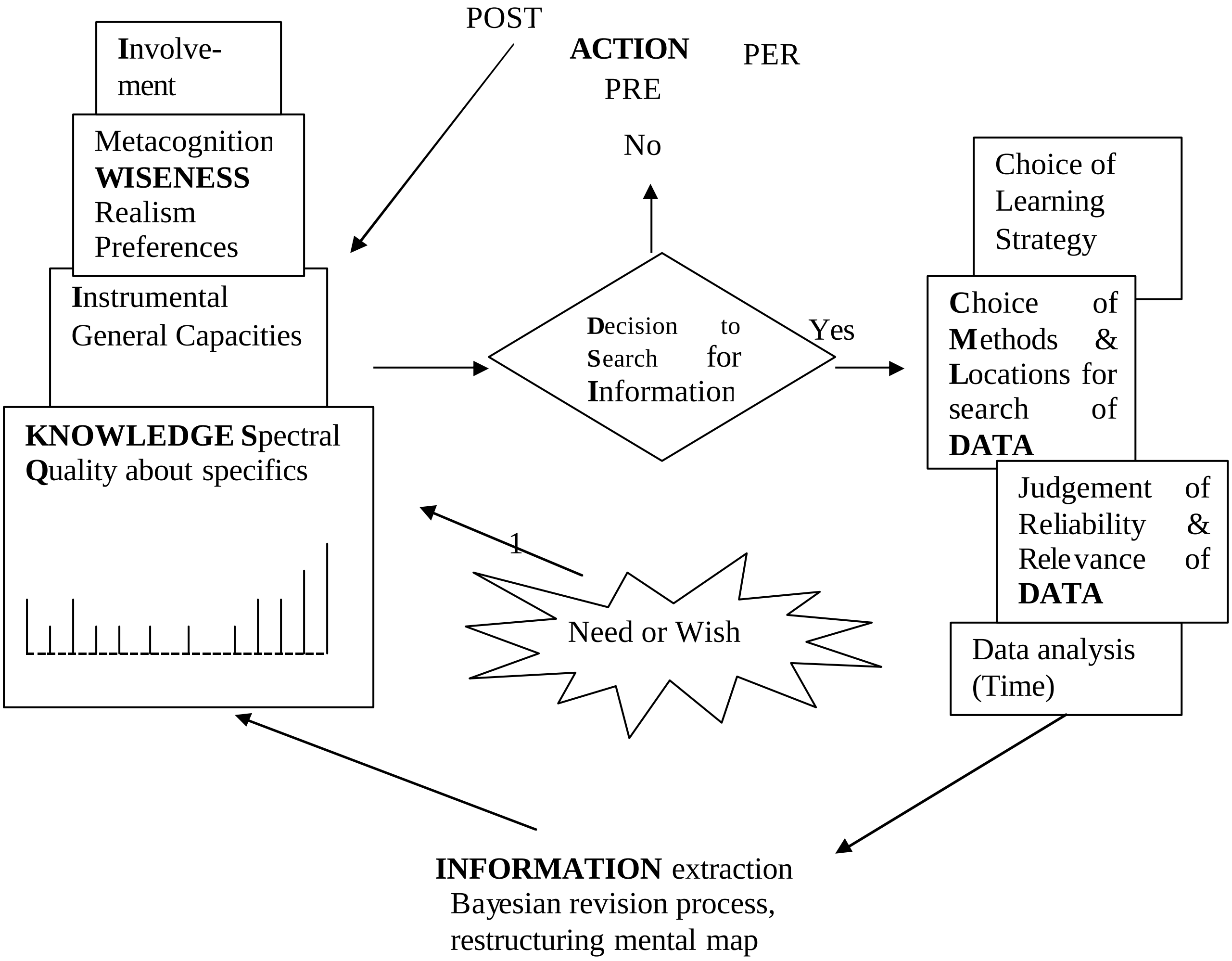
(LS) Judgement of relevance of data : Leclercq & Pierret (1989,) demonstrated that students select the type of data they request according to their learning needs. In their experiment, students had to study two contents presented in a hypermedia format. For one content, they were told that the exam would be a MCQ test ; they asked more frequently to be tested on content by MCQ. For the other content, they were told that the exam would be an oral presentation ; they asked more frequently for synthetic version of the content.

(CM) Revision of subjective probabilities about an event. Leclercq (1983, 270) have designed laboratory experiments based on a modified version of the "Shannon Guessing Game" (Attneave, 1975), i.e. guessing the successive letters of words in a text, before and after receiving clues. They have demonstrated that students process information according to Bayes'theorem of revision of probabilities in most cases, but not in all cases, and they have suggested explanations for these differences.

(CM)-<u>observation of subtle changes</u> in mental structures or states. Leclercq & al. (1999) have designed a situation where students were presented numerous situations of class disruption by pupils. For each situation, student-teachers had to make a choice (with a confidence degree) between different courses of (re)action. Then a debate took place and was followed by choosing again among the same courses of action. It appeared that students did not change their choices, but they changed their confidence degrees (see a typical result in the graph hereafter). The students realising that the majority had made the same choice as their one increased their confidence (except the student represented by the big circle). The students of the minority opinion did not change their response, but lowered their confidence degree (in this example, only one, represented by the big square).



The following schema[4] shows the SPIRAL move of learning under the impulse of a need or a wish (starting point) :



Conclusion

Everybody acknowledges the importance of metacognitive skills and mathetic[5] competencies for nowadays learners. Self assessment is only an aspect of them and Confidence Degrees 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 Confidence Degrees, 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

[4] The representation of the left is Leclercq's pyramid of competencies (1998,52).

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

confident that this approach can bring its special light in the old debate so well stated by an Arab proverb[6] :

"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

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.

Bruttomesso*, D, R. Gagnayre•, D. Crazzolara*, , E. Busata*, D.Leclercq°, J.F.d'Ivernois•, E. Casiglia*, A. Tiengo*, A. Baritussio#. Confidence degrees to evaluate knowledge in patients with Type 1 diabetes, Patient Education and Counseling, paper accepted, 2002

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.

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.

Leclercq, D. (1983). 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. 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.

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.

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

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