

## A software for a new data analysis method : CHIC

Marc BAILLEUL

Institut Universitaire de Formation des Maîtres  
CAEN, FRANCE

### Abstract :

CHIC : what does it stand for ?

Hierarchical Implicative and Cohesive Classification

What kind of data does it work with ?

- Numeric data, binary or not but with values between 0 and 1, presented in a table, for example 3 000 students x 250 variables

- Situations of presence / absence, satisfaction degrees, ...

What can we do with it ?

Particular Implicative Statistical Analysis, developed by Régis GRAS and his students. This method is used to show hierarchised nets of variables. It has been used in numerous studies in didactic of mathematics in France.

## A software for a new data analysis method : CHIC

This communication will be composed of two parts :

- 1) a short one about the mathematical foundation of this software,
- 2) a demonstration of the principal functions. We shall show these functions through the analysis of a questionnaire about problem solving in a group of primary teachers.

### Implicative statistical analysis

Every researcher interested in the relations between variables (for example a psychologist, a specialist of methods, a didactic specialist...) asks himself as follows : « Let  $a$  and  $b$  be two binary variables, can I affirm that the observation of  $a$  leads to the observation of  $b$  ? ». In fact, this non-symmetrical point of view on the couple  $(a, b)$ , unlike of the methods of similarity analysis, expresses itself by the question : « Is it right that if  $a$  then  $b$  ? ». Generally, the strict answer is not possible and the researcher must content himself with a quasi-implication. We propose, with the statistical implication, a concept and a method which allows us to measure the degree of validity of an implicative proposition between (binary or not) variables. Furthermore, this method of data analysis allows us to represent the partial order (or pre-order) which structures a set of variables.

### Theoretical aspects of the binary case

Let us consider the generic situation of the binary case. We cross a set  $E$  of objects and a set  $V$  of variables. We now want to give a statistical meaning to a quasi-implication  $a \Rightarrow b$  (logical implications are exceptional). We note  $A$  (respectively  $B$ ) the subset of  $E$  where the variable  $a$  (respectively  $b$ ) takes the value 1 (or true). Measuring the quasi-inclusion of  $A$  into  $B$  is similar to measuring this reduced form of implication. Intuitively and qualitatively, we can say that  $a \Rightarrow b$  is admissible if the number of counter-example (objects of  $A \cap \bar{B}$ ), verifying  $a \wedge \bar{b}$  in  $E$  is improbably small compared to the number of objects expected in an absence-of-a-link hypothesis between  $a$  and  $b$  (or  $A$  and  $B$ ).

The quality of the implication is measured with the implication intensity. The approach developed for elaborating the implication intensity is inspired by I. C. Lerman's theoretical considerations for designing his similarity indexes (Lerman, 1981). We associate  $A$  (and  $B$  respectively) with a random subset  $X$  (and respectively  $Y$ ) of  $E$  which have the same cardinal. We then compare the cardinal of  $A \cap \bar{B}$  with the one of  $X \cap \bar{Y}$  in an absence-of-a-link hypothesis. If the cardinal of  $A \cap \bar{B}$  is

improbably small in comparison with the cardinal of  $X \cap \bar{Y}$ , the quasi-implication  $a \Rightarrow b$  will be accepted; otherwise, it will be refused. It has been demonstrated (Lerman et al., 1981), that the random variable  $\text{card}(X \cap \bar{Y})$  follows a hypergeometrical law and, under certain conditions, follows a Poisson's law of parameter  $\text{card}(A) \cdot \text{card}(\bar{B}) / \text{card}(E)$ . The implication intensity is defined by the function :  $\mathbf{j}(a, \bar{b}) = 1 - \Pr(\text{card}(X \cap \bar{Y}) \leq \text{card}(A \cap \bar{B})) = 1 - \sum_{i=0}^{\text{card}(A \cap \bar{B})} \frac{\mathbf{I}^i}{i!} e^{-\mathbf{I}}$ ,

with  $\mathbf{I} = \text{card}(A) \cdot \text{card}(\bar{B}) / \text{card}(E)$ . We can say that the quasi-implication  $a \Rightarrow b$  is admissible at the level of confidence  $\mathbf{a}$  if and only if  $\mathbf{j}(a, \bar{b}) \geq 1 - \mathbf{a}$ .

For example, we have 100 students ( $\text{card}(E) = 100$ ) who can have two behaviours a and b with  $\text{card}(A) = 6$ ,  $\text{card}(B) = 15$  and  $\text{card}(A \cap \bar{B}) = 1$ . We can observe that the number of students (here 1) refuting the implication  $a \Rightarrow b$  is improbably small in an absence-of-a-link hypothesis. In fact,  $\mathbf{j}(a, \bar{b}) = 0.965$  that is to say a level of confidence equal to 96.5 per cent for the implication because the probability that  $\text{card}(X \cap \bar{Y}) < 1$  is equal to 0.035.

The notion of statistical implication is extended :

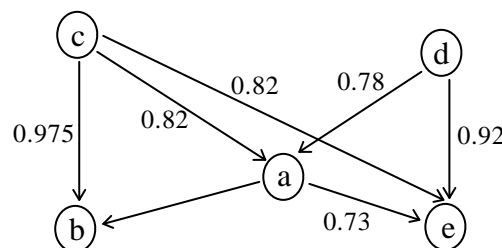
- to modal (or qualitative) variables and numerical (or quantitative) variables in (Larher, 1991),
- to ordinal variables in (Bailleul, 1994).

### Implication graph

A great interest of the statistical implication consists in studying together all the variables on a given population. We can associate a measure of their implication intensity with each couple  $(a, b)$  of variables. This will be represented by a valued oriented edge. When the cardinals of A and B are equals, there are two oriented edges ( $a \rightarrow b$  and  $b \rightarrow a$ ). If we fix a condition of transitivity of the implication (generally 0.5), it is possible to generate a transitive graph. For example, if we have a set of five variables, whose implication intensities greater than 0.5 are given in the following table :

$\Rightarrow$	a	b	c	d	e
$\Rightarrow$	a	b	c	d	e
a		0.97			0.73
b					
c	0.82	0.975			0.82
d	0.78				0.92
e					

we obtain the following graph :



A. Larher (1991) has proved that the order between the intensities respects the order between the cardinals. So, for each pair of variables, we only keep the maximal intensity of the two couples

defined by this pair. We can also prove (Gras et Larher, 1992) the relation existing between the linear correlation coefficient and the statistical implication and the relation between the  $\chi^2$  of independance and the statistical implication.

M. Bailleul (1994) has built the notion of subjects' contribution to the transitive ways and nets of variables. So it is possible to find subjects who are representative of each net of variables.

## Representation of problem solving for elementary school teachers

### Methodology

In 94-95, we proposed a questionnaire to 97 elementary school teachers and asked them to choose in this set of assertions five sentences they agreed with and five sentences they disagreed with. They had also to order these sentences from 1 (resp. (-1)) to 5 (resp. (-5)), 5 was given to the sentence with which they agreed most (resp. disagreement).

Here is the questionnaire.

n°	assertion	«mark»
1	In mathematics, what pupils of my class prefer is to solve problems.	
2	Before letting pupils work on a problem, we solve one together at the board.	
3	In a group work, when the first ideas are expressed, I like the research to become individual.	
4	When a problem is given, whatever may be the way of working, we must obtain a common solution.	
5	I often give problems about new notions.	
6	Problem solving takes a lot of time. If we want to cover the curriculum entirely, it would be better to give pupils training exercices.	
7	Generally, I give the pupils a problem and they get along.	
8	I don't let pupils write mistakes on the board so as not to distort the ideas.	
9	We often bore the pupils with problems they are not interested in.	
10	When I propose problem-solving, I'm thinking of a « working-group ».	
11	In the first primary grade, I think pupils don't have to solve problems.	
12	Even if it takes a lot of time, I let pupils solve problems with their own capacities.	
13	I mark the problems that I have given to the pupils.	
14	When we correct a problem, every pupil is invited to propose his solution to the other pupils.	
15	I prefer giving a problem when a notion is completely explained. It prevents pupils from making mistakes.	
16	Generally, pupils enjoy doing problems.	
17	I reserve problem solving for the end of the school year.	
18	Problems are, for me, the occasion of seeing if pupils have a synthesis view of what they have been studying so far.	
19	I do not mind staying more than a week on a problem if I see that pupils keep trying to find the solution.	
20	When we correct an exercice, I eventually let a pupil write a wrong proceeding on the board so that a debate may occur, before unanimity.	
21	If we propose many researches, pupils don't like easy exercices ; they prefer « brain-breaking » problems.	
22	I don't let pupils go wrong because it makes us lose time.	

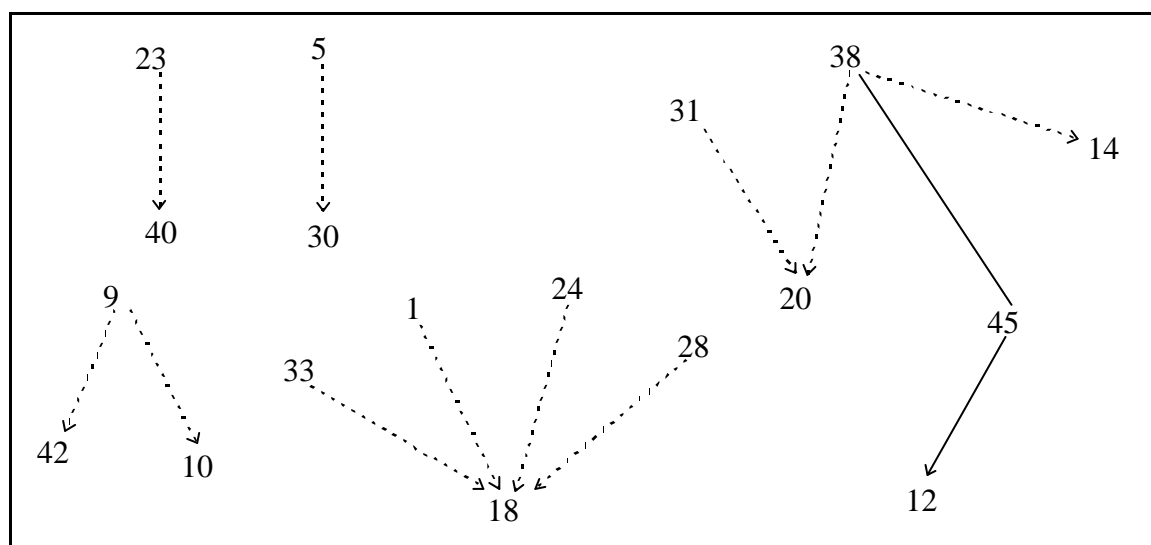
We have obtained a table (97 lines and 23 columns) where variables were ordinal values. We transformed it in a 97 x 46 table as follows :

- the first 23 columns correspond to the 23 sentences « positively » seen : someone chooses it and adheres to this assertion giving it a positive row. Ordinal variables become modal variables like this : not chosen  $\rightarrow 0$  ; choice 1  $\rightarrow 0.2$  ; choice 2  $\rightarrow 0.4$  ; choice 3  $\rightarrow 0.6$  ; choice 4  $\rightarrow 0.8$  ; choice 5  $\rightarrow 1$ .
- the next 23 columns correspond to the 23 sentences « negatively » seen : someone chooses it and reject this assertion giving it a negative row. Ordinal variables become modal variables like this : not chosen  $\rightarrow 0$  ; choice (-1)  $\rightarrow 0.2$  ; choice (-2)  $\rightarrow 0.4$  ; choice (-3)  $\rightarrow 0.6$  ; choice (-4)  $\rightarrow 0.8$  ; choice (-5)  $\rightarrow 1$ .

We shall find in the next table the set of variables and their weight. (si = sentence n°i)

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12
	s1+	s2+	s3+	s4+	s5+	s6+	s7+	s8+	s9+	s10+	s11+	s12+
weight	2.8	7.6	29.2	8	18.6	1.4	14.4	3.4	5.6	7.8	0	34
	v13	v14	v15	v16	v17	v18	v19	v20	v21	v22	v23	
	s13+	s14+	s15+	s16+	s17+	s18+	s19+	s20+	s21+	s22+	s23+	
weight	1	49	2.8	7.6	0	23.2	8.2	51.2	1.6	5.6	5.4	
	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	v35
	s1-	s2-	s3-	s4-	s5-	s6-	s7-	s8-	s9-	s10-	s11-	s12-
weight	5.4	18	1	17.2	14.4	27	9.2	26.8	6	7.4	30	2.4
	v36	v37	v38	v39	v40	v41	v42	v43	v44	v45	v46	
	s13-	s14-	s15-	s16-	s17-	s18-	s19-	s20-	s21-	s22-	s23-	
weight	16.6	0.8	14.6	2.4	53.2	0.4	10.2	1.8	3	20.2	3	

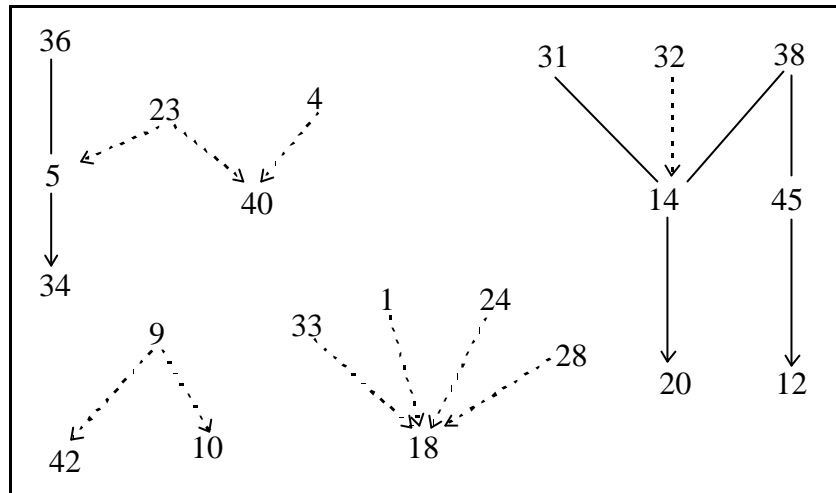
Here, we shall be specially interested in the way variables nets appear in the implicative graph.



### Implicative graph built at threshold .75

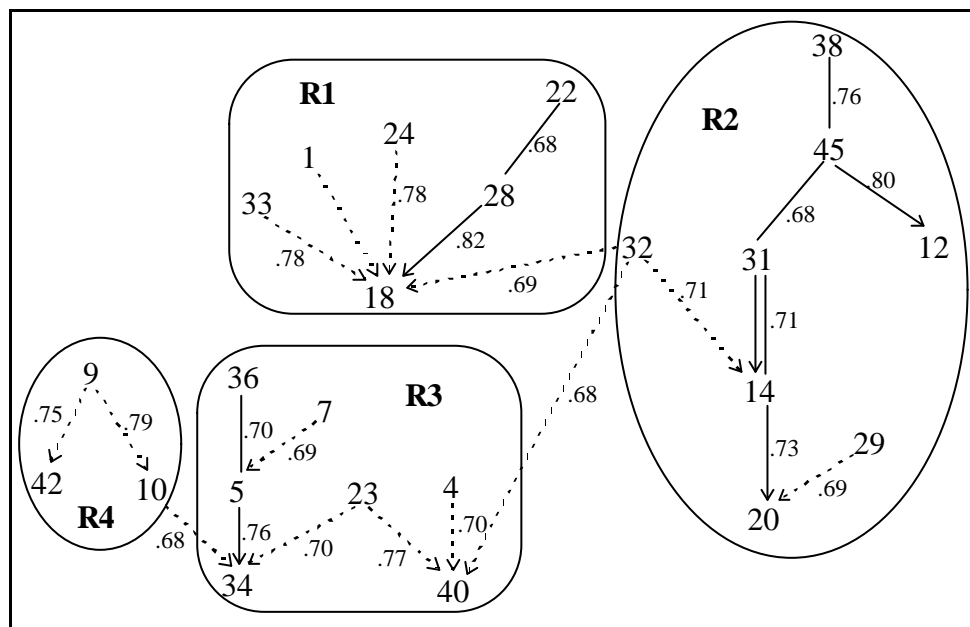
In this graph, thick lines represent transitive links.

We shall progressively lower the threshold of the intensity of implication to try to perceive the complexity in the set of implicative links between variables.



### Implicative graph built at threshold .70

We can clearly distinguish four nets of variables. Let us lower a little more the threshold.



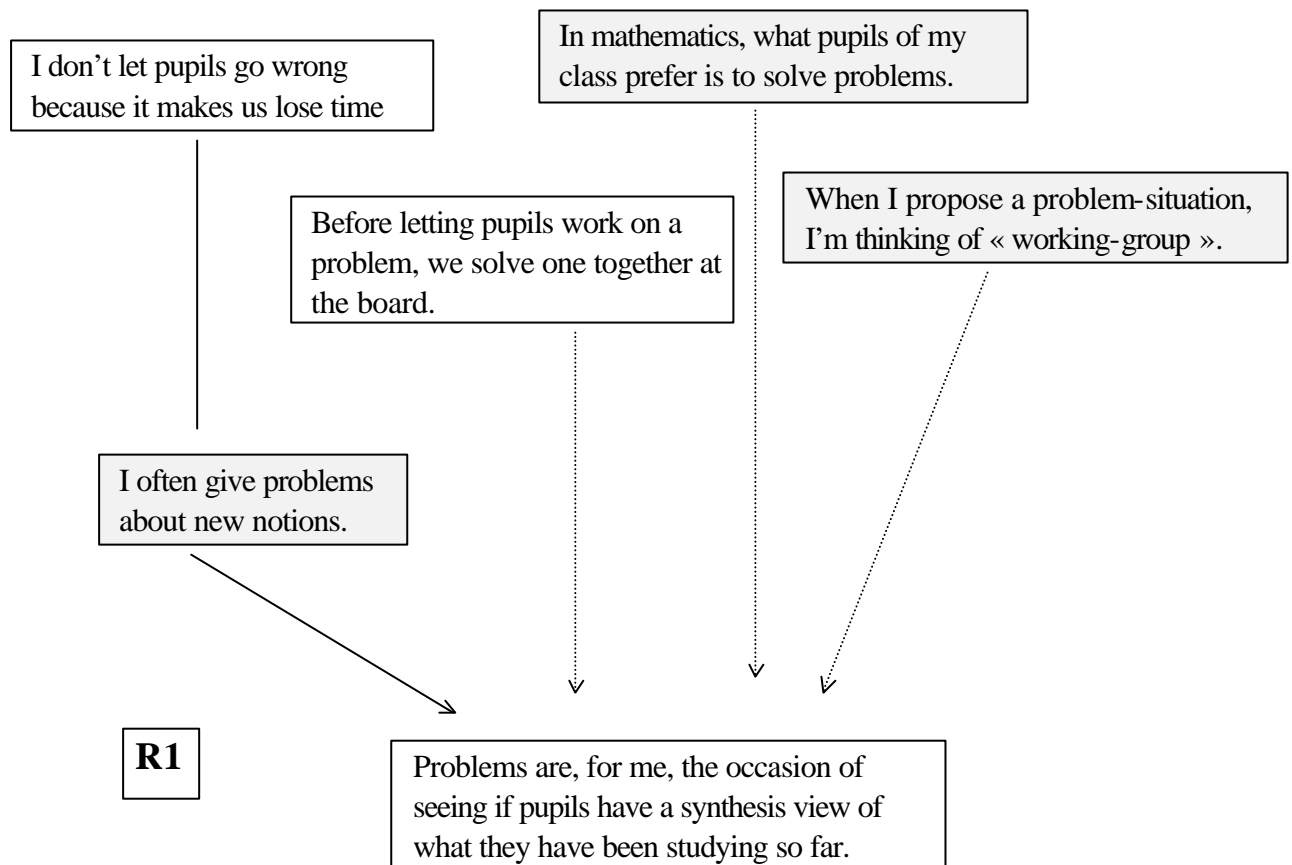
### Implicative graph built at threshold .68

We stop here our investigation of the implicative graph because of a risk : if we lower more the threshold, the quality of links does not remain as sure as we want.

This analysis has shown four nets of variables that we now have to translate (write the assertions that are behind the numbers of the variables) to give interpretation of the nets.

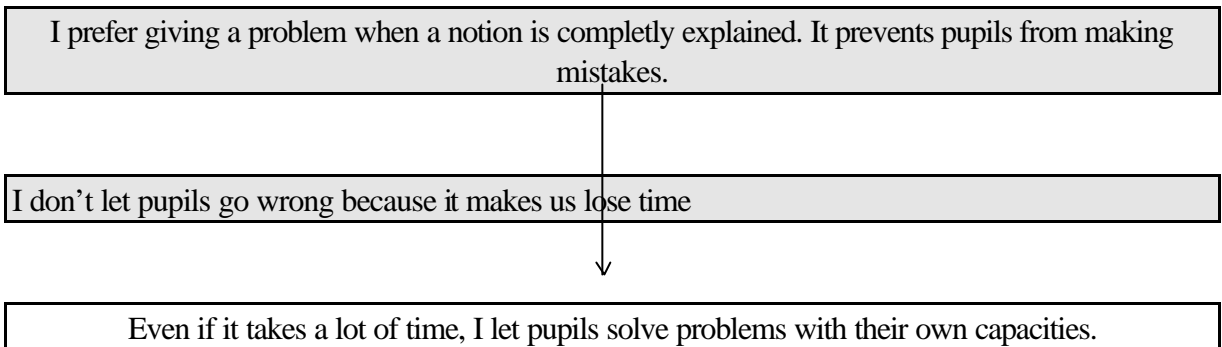
The result of the first operation is shown in the following page.  
What do these four nets of variables mean ?

Let us transcribe sentences of R1. In the diagrams below, a frame with a white background means a positive opinion on the assertion, a frame with a grey background means a negative opinion on the assertion.



Through this organisation, it seems that we have here a representation of « evaluation-problem », a synthesis during which pupils, individually, have to reproduce some situations that have already been proposed.

Let us now study the implications which constitute the net R2.



We can see in this transitive way the choice to give pupils problem as tools to build notions, even if we risk being faced with difficulties they will have to try to overcome with their own capacities. This choice required some « time-investment ».

I don't let pupils go wrong because it makes us lose time.

I don't let pupils write mistakes on the board not to distort the ideas.



When we correct a problem, every pupil is invited to propose his solution to the other pupils.

The risk of deadlocks and the risk of «distorting the ideas » are taken upon themselves by the teacher through a time of confrontation in the group of pupils that have, here, a regulating cognitive function.

I don't let pupils write mistakes on the board not to distort the ideas.

When we correct a problem, every pupil is invited to propose his solution to the other pupils.



When we correct an exercise, I can let a pupil write a wrong proceeding on the board so that a debate may occur, before unanimity.

We can state precisely the previous interpretation : the regulating cognitive function of the pupils' group is at the origin of unanimity.

There are also two simple implication in R2.

We often bore the pupils with problems they are not interested in.



When we correct a problem, every pupil is invited to propose his solution to the other pupils.

We propose this interpretation : one way of getting pupils interested in solving problems is to make of this activity a « social » activity in the group of pupils.

Problem solving takes a lot of time. If we want to cover the curriculum entirely, it would be better to give pupils training exercises.

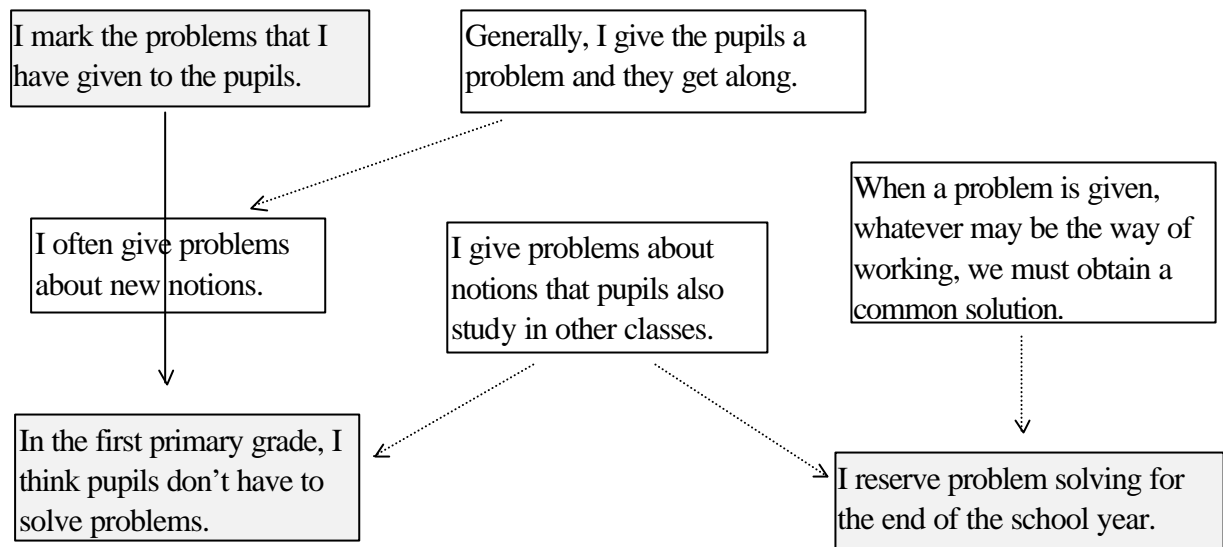


When we correct an exercise, I can let a pupil write a wrong proceeding on the board so that a debate may occur, before unanimity.

We can see here the same idea of a « social » activity reinforced by the negative opinion on the training exercises.

To sum up R2, we shall say that it shows us an obvious evolution of the function of error. Errors, and the capacity to go beyond, are for a pupil, a singular individual in a group, constituents of the sense of problem-solving.

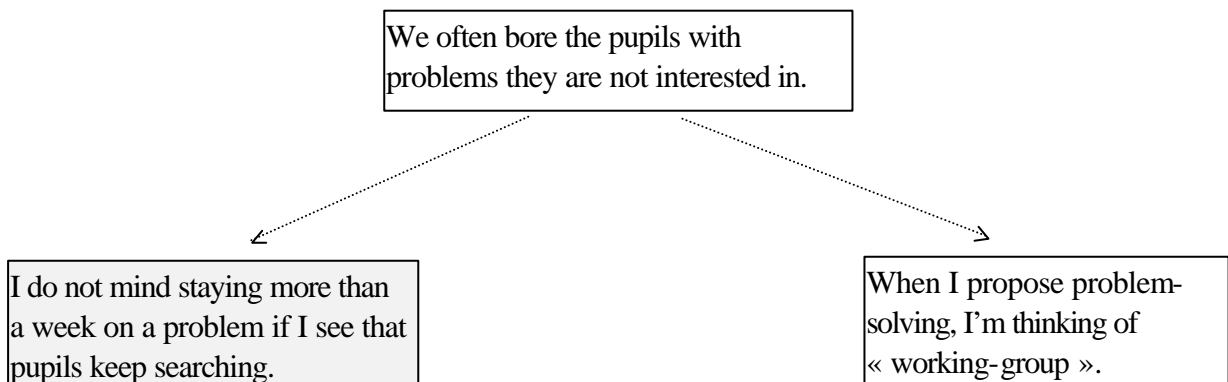
The following net R3 is more difficult to interpret than the two previous ones. Let us transcribe it.



We think that in this net we can see problem-solving as a tool for global learning:

- learning of self-governing : pupils get along in front of a new situation, even about new notions;
- learning of socialisation, and cognitive socialisation : we must obtain a common solution. This preoccupation is a purpose all through the year, problems are not reserved for the end of the year;
- learning mathematics with subjects of other classes.

The last net R4, composed of only two implications, is teacher-centred.



This net shows a pedagogical problem for the teacher: the fears that problems may bore pupils. The solution is « working-group », but it requires a lot of time!

We are now able to sum up the characteristics of the four nets that have come to light thanks to the implicative analysis:

**R1: evaluation-problem;**

**R2: errors, and the capacity to go beyond, as constituents of sense of problem-solving;**

**R3: problem-solving as a tool for global learning;**

**R4: problem-solving as a pedagogical problem for the teacher.**