
Fonctionnement didactique de la simulation en statistique

Exemple de l'enseignement du concept de corrélation

Jean-Claude Oriol* — **Jean-Claude Régnier****

**Institut Universitaire de Technologie Lumière
Université Lumière, Lyon II F-69676 BRON Cedex
ÉCOLE Doctorale ECLIPS (UNIVERSITE LYON 2)
Centre d'Études, de Recherche et de Recherche-Action (CERRAL)
Jean-Claude.Oriol@univ-lyon2.fr*

*** Département des Sciences de l'éducation
Université Lumière Lyon II 16 Quai Claude Bernard F69007 Lyon
Laboratoire Interdisciplinaire de Recherche sur les Curricula et les Apprentissages
Jean-Claude.regnier@univ-lyon2.fr*

1. Introduction

Notre propos s'ancre sur la prise en compte de la place de la simulation dans la pratique de l'enseignement de la statistique et dans le champ théorique de la didactique de la statistique. La simulation ressort dans nombre d'ouvrages de référence, de directives officielles, de commentaires des programmes comme un outil privilégié qui doit être mis au service des apprenants dans les situations didactiques visant des notions du champ de la statistique. Ainsi de nombreux logiciels de simulation plus ou moins sophistiqués, élaborés à l'intention des étudiants et des enseignants sont apparus. Cette panoplie va des langages dédiés jusqu'aux outils généralistes. À ce jour le langage Java permet de produire des Applets qui offrent des possibilités considérables. Et même si notre propos n'est pas de développer ici cette idée, nous pensons que cet intérêt pour les simulations et cette insistance à les intégrer dans des séquences didactiques concernant les apprentissages statistiques sont d'une certaine manière significatifs d'une spécificité endogène au champ de la statistique.

2. Quelques distinctions concernant les simulations utilisées en statistique

Comme le signale Jamie D. Mills (Mills, 2002) on peut distinguer dans la littérature quatre manières d'envisager l'intégration de simulations dans l'apprentissage de la statistique : *" Four definitions of computer simulation methods were described in the literature reviewed for this article. One definition involved students writing their own programs (using SAS PROC IML®, say), setting up a model for a problem and investigating diagnostics for the model in seeking possible violations of assumptions. A second definition allowed students to experience similar advantages using a random number generator in Excel® or MINITAB®. Using Excel® or MINITAB®, the commands to generate the random samples and perform experiments on the model are mostly window-driven. Third, many instructors used some combination of the first and second definitions, by providing program templates that allowed students to change parameters during the experiments (commonly in SAS® or SPSS®). Finally, a fourth definition involved using commercial software packages designed exclusively for simulation purposes (for example, the "Samplings Distribution" program). The literature reviewed in this paper included journal articles that utilized all four operational definitions, although the majority of the authors reported using the latter three. "*

3. Le contexte de l'IUT Lumière

3.1. L'alternance à l'IUT Lumière

Depuis sa création en 1992 l'IUT Lumière est le seul IUT dans lequel la formation des étudiants est complètement réalisée en alternance. Cette formation est ainsi construite : en première année les étudiants suivent un stage de sept semaines dans une entreprise pendant la période avril-mai, la deuxième année est réellement en alternance sur la base d'un contrat d'apprentissage. Deux modalités sont en place en GEA selon l'option choisie en deuxième année par les étudiants 2 jours + 3 jours, 7 semaines + 7 semaines, tandis qu'une autre gestion

du temps, à savoir 15 jours à l'IUT + 15 jours en entreprise, est organisée dans les trois autres départements. Dans tous les cas, l'année universitaire s'achève par la tenue des jurys et la délivrance des diplômes au mois de septembre. Cette forme d'organisation des études est usuellement désignée par «système 1+1» . On trouvera des compléments détaillés sur l'organisation des cursus dans une communication de Paul Rousset, directeur de l'IUT depuis sa création (Rousset, 1997), et plus particulièrement sur le département GEA dans un article d'Isabelle Barth (Barth, 2001) qui a dirigé le département GEA pendant plusieurs années.

3.2. L'organisation de l'enseignement de la statistique dans le département GEA

Le temps passé en entreprise est modulable entre la première et la deuxième année. Il en résulte que l'équipe enseignante a pondéré différemment les cours de statistique dans chacune des deux années. En première année, une quarantaine d'heures est consacrée à l'étude de la statistique descriptive. Cet enseignement donne lieu à deux contrôles sous la forme d'épreuves individuelles en temps limité et à une évaluation à partir d'une enquête impliquant plusieurs disciplines (communication, marketing, statistique, etc.) et réalisée en groupes. En deuxième année, seize heures sont consacrées à travailler à l'étude de la statistique inférentielle, en particulier sur les notions de lois de probabilité, d'échantillonnage, d'estimation, d'intervalle de confiance

4. Le cas particulier du concept de corrélation

4.1 le cadre théorique

Le point de vue de la formation en alternance nous renvoie à une question fondamentale : Comment apprenons-nous ?

La théorie des champs conceptuels de Gérard Vergnaud (Vergnaud 1991 p.135) répond de manière satisfaisante pour ce qui touche aux apprentissages en statistique : « *La théorie des*

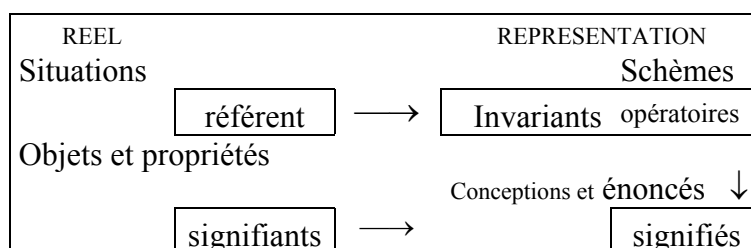
champs conceptuels est une théorie cognitiviste, qui vise à fournir un cadre cohérent et quelques principes de base pour l'étude du développement et l'apprentissage des compétences complexes, notamment celles qui relèvent des sciences et des techniques.(...) [et] pour la compréhension des filiations et des ruptures entre connaissances ».

La conceptualisation, élément central de l'apprentissage, prend appui sur le concept conçu comme « un triplet de trois ensembles référence, signifié et signifiant :

- *l'ensemble des situations qui donnent du sens au concept (la référence),*
- *l'ensemble des invariants sur lesquels repose l'opérationnalité des schèmes (le signifié)*
- *l'ensemble des formes langagières et non langagières qui permettent de représenter symboliquement le concept, ses propriétés, les procédures de traitement (le signifiant) »*

Suivons Gérard Vergnaud (Vergnaud 1994, p. 180) qui pour affiner « *progressivement la définition d'un schème,..[dit]... d'abord que c'est une totalité dynamique fonctionnelle, c'est-à-dire quelque chose qui fonctionne comme une unité ; en second lieu que c'est une organisation invariante de la conduite pour une classe de situations données (l'algorithme est un cas particulier du schème) ; et en troisième lieu qu'un schème est composé de quatre catégories d'éléments : 1)des buts, intentions et anticipations ; 2)des règles d'action; 3)des invariants opératoires; 4)des possibilités d'inférences en situation. »*

Revenant sur ce qu'il appelle lui-même le « triangle des linguistes », Vergnaud (Vergnaud, 1994, p. 189-190) propose le modèle suivant :



Ainsi la notion de champ conceptuel en tant qu'ensemble des situations renvoyant à l'idée de procédure, permet de situer la simulation en statistique dans sa finalité qui est de fournir à l'apprenant un éclairage sur le signifié (invariants opératoires)

Par ailleurs l'approche développée par Jean-Claude Régnier (Régnier 1988) intégrant l'apprentissage fondé sur le tâtonnement expérimental de l'apprenant offre une perspective pour analyse didactique de la simulation et relation à la résolution de problème.

4.2 Quelques travaux concernant la corrélation et la simulation

Mais du point de vue général sur l'apprentissage revenons à celui, plus particulier, de l'apprentissage de la corrélation passant par la simulation. La plupart des travaux publiés sur ce sujet sont américains. Citons brièvement Franklin (1992) qui propose l'utilisation de Minitab®, mais également Jensen (1983), Pulley et Doolbear (1984), Olinsky et Schumacher (1990), Ferral (1995), Romeu(1995) qui parmi d'autres proposent des approches des coefficients de régression et de détermination utilisant des générateurs de nombres aléatoires. Batanero et Godino (2001 page 4-15) proposent des études de la corrélation utilisant STATGRAPHICS ; ils citent Inhelder et Piaget ces derniers relevant « *que algunos sujetos analizan solamente la relacion entre los casos favorables positivos* », et ils font également référence à Chapman y Chapman (1967) qui développent la notion de « *contingencia empiricas* ».

5. La situation proprement dite

5.1 Le choix d'Excel

Notre pratique pédagogique vise à intégrer le plus tôt possible l'outil informatique comme instrument canonique d'une pratique de la statistique. Et en matière d'outil informatique, le

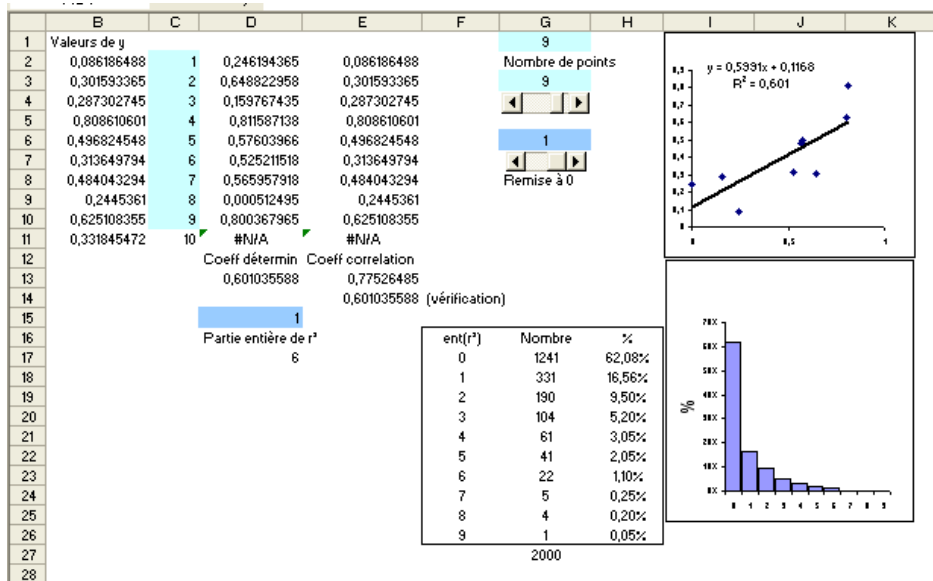
tableur est privilégié. Plusieurs raisons expliquent ce choix surprenant en comparaison de la puissance de certains logiciels de traitements statistiques. Les cours concernent des étudiants de premier cycle " généralistes ". Le tableur est dans ce cadre utilisé par différentes disciplines (comptabilité, gestion, mathématiques, mathématiques financières) et l'investissement sur un autre logiciel serait peu rentable.

5.2 La situation problème et la construction de la simulation

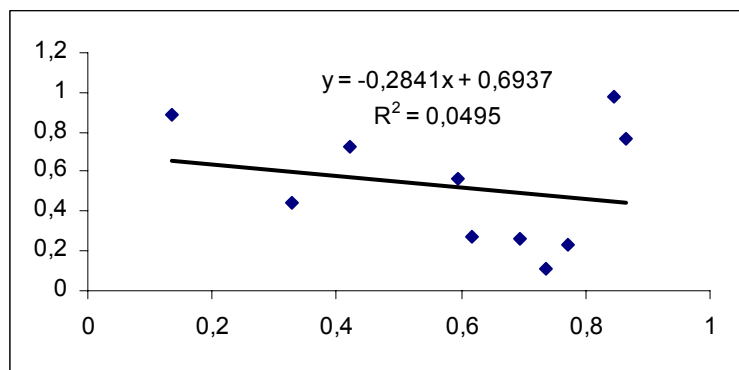
La séance didactique prend appui sur une bonne maîtrise du logiciel Excel par les étudiants et fait suite à une séance de quatre heures sur la corrélation. Les consignes sont les suivantes :
« *Vous devez construire un classeur Excel, sans utiliser le langage VBA, qui contienne les opérations permettant de tirer au hasard $2N$ nombres compris entre 0 et 1, de représenter sur un graphique N couples ayant pour coordonnées les nombres précédents, de calculer le coefficient de détermination.*

Vous devez conserver les résultats dans un tableau et représenter ce tableau par un graphique. Vous devez construire un dispositif de remise à 0 et un autre dispositif permettant de choisir le nombre N de points. »

Par binôme, les étudiants produisent un classeur dont la feuille de visualisation a, aux variations individuelles près, l'aspect suivant :



La représentation graphique n'appelle que peu de commentaires :



Dans le tableau : Première colonne : valeur de R^2 première ligne entre 0 et 0,1 deuxième ligne entre 0,1 et 0,2 etc.

Deuxième colonne : nombre de configurations

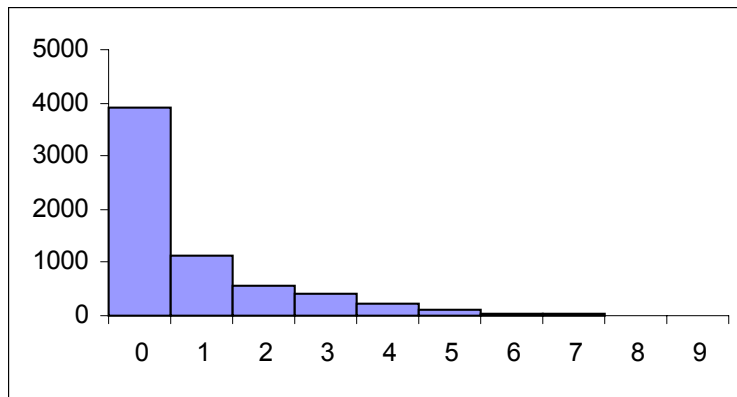
Troisième colonne : pourcentage sur le total

0	3846	60,63%
1	1113	17,55%
2	573	9,03%
3	395	6,23%
4	223	3,52%
5	112	1,77%
6	51	0,80%
7	25	0,39%
8	6	0,09%
9	0	0,00%

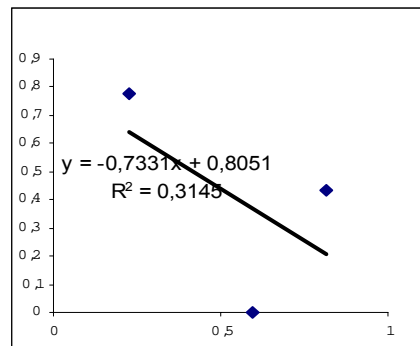
6344

Et enfin un graphique représentant le tableau précédent

Pour dix points et 6344 essais :



Voici d'autres résultats pour trois points :



On en déduit le tableau suivant agrémenté du même commentaire

Première colonne : valeur de R^2 première ligne entre 0 et 0,1 deuxième ligne entre 0,1 et 0,2 etc

Deuxième colonne : nombre de configurations

Troisième colonne : pourcentage sur le total

ent(r^2)	Nombre	%
0	2064	20,64%
1	935	9,35%
2	720	7,20%
3	681	6,81%
4	648	6,48%
5	672	6,72%
6	713	7,13%
7	737	7,37%
8	844	8,44%
9	1986	19,86%

10000

6. Retour sur la situation avec des prolongements éventuels

Cette situation a le mérite de poser très fortement le problème de la validité du coefficient de corrélation en fonction du nombre de points envisagé et de trouver des solutions.

En étudiant les cours de statistique on remarque en effet que c'est dans les cours où l'on peut le moins répondre en termes d'intervalles de confiance à cette question que l'on s'en sert le plus abondamment !

Dans l'exposé, nous présenterons des solutions apportées aux étudiants par la lecture d'abaques permettant de répondre à cette interrogation, pensant que l'important et plus de développer un questionnement et une interrogation sur les données et sur les résultats des traitements qu'elles subissent, plutôt que des savoirs morts attachés trop souvent aux techniques probabilistes sophistiquées.

Nous pourrions alors, à l'aide de la théorie des situations didactiques (Brousseau 1998) développée par Guy Brousseau essayer d'aborder ce qu'il appelle les obstacles macro-didactiques (Brousseau 2003) :

« L'intégration du milieu, y compris le milieu socio-culturel et professionnel des enseignants et des élèves dans les objets de la théorie des situations permet aujourd'hui d'identifier d'autres types d'obstacles dits « macro-didactiques ». Ces obstacles à la diffusion d'une connaissance proviennent des rapports des diverses institutions d'une société à ce savoir. En statistique notamment de nombreux auteurs (par exemple Daniel SCHWARTZ) ont noté que la Grande Bretagne et les Etats-Unis bénéficiaient de conditions socioculturelles beaucoup plus favorables à cet enseignement que la France. Les difficultés proviendraient des représentations (au sens de la sociologie) que se font ces institutions de la façon de pratiquer, d'apprendre ou d'enseigner ce savoir en le reliant aux mathématiques (esprit cartésien, individualisme, méconnaissance de la variance etc.), qui entraîneraient des difficultés récurrentes. Malgré l'existence de solutions microdidactiques et la possibilité évidente de les mettre en œuvre, le développement ne se ferait pas pour des raisons de ce genre. Le phénomène s'observe dans différents secteurs des mathématiques mais chaque fois de manière spécifique à un certain savoir. Ce type d'obstacles est particulièrement évident en statistique et demanderait une étude qui sort du cadre de ce cours »

Nous prolongerons également l'étude par un exposé sur le problème particulier posé par trois points et des résultats « étranges » dans ce cas.

7. Un « renversement de paradigme » : la coopération entre groupes

Une autre situation didactique consiste à donner aux étudiants la consigne de finaliser le produit (ici le classeur Excel) afin qu'il puisse servir dans d'autres contextes :

- Démonstration aux pairs.
- Utilisation par d'autres groupes avec échanges, par exemple : groupe A échange simulation A (sur le coefficient de variation), contre la simulation B (sur les intervalles de confiance) faite par le groupe B.
- Utilisation dans un contexte d'enquête : voir à ce propos (Oriol, 2002)
- Travail avec des groupes éloignés : le produit est destiné à un public ciblé en âge et en termes de compétences avec retour et dialogues par internet.

8. Conclusion

Nous ne serons pas ici les thuriféraires d'une doxa défendant un point de vue ou un logiciel particulier, mais plutôt dans une approche d'ingénierie didactique dans laquelle la simulation joue un rôle dans la fragile catalyse qui accompagne la construction des savoirs. Ici, la construction par les apprenants, eux-mêmes, de l'outil s'apparente aux méthodes d'apprentissage où les premiers travaux des apprentis consistaient à fabriquer leurs propres outils (l'équerre de l'outilier par exemple) tout en mettant en scène le tâtonnement expérimental producteur d'apprentissages. Parmi les problèmes en suspens reste le contrôle par les étudiants du générateur de nombres aléatoires d'Excel. A la lumière de la théorie des champs conceptuels et du modèle d'apprentissage fondé sur le tâtonnement expérimental de l'apprenant, nous présenterons nos résultats issus de l'expérimentation et de l'observation, en particulier ceux liés à l'effet des variables d'énoncé du problème sur l'activité des étudiants et sur leur performances. Nous prolongerons nos hypothèses par l'observation de la façon dont cette activité de simulation se contextualise dans la pratique d'enquêtes par questionnaires réalisées par les étudiants dans la suite de cet enseignement.

9. Bibliographie

- Albert, J. H. (1993), "Teaching Bayesian Statistics Using Sampling Methods and MINITAB," *The American Statistician*, 47, 182-191.
- Antoine F., Grootaers D., Tilman F., (1988), *Manuel de la formation en alternance*, Lyon, Chronique sociale : Bruxelles, Vie ouvrière.
- Bachelard G., (1938), *La formation de l'esprit scientifique*, Paris Vrin.
- Barth I., (2001), « L'utilisation de la situation de travail dans la formation d'un cursus diplômant, cas du DUT GEA de l'IUT Lumière », *Deuxième rencontre des acteurs de la formation technologique d'Europe et d'Amérique Latine*, Pointe à Pitre.
- Batareno C. et Godino J., (2001), *Análisis de Datos y su Didáctica*, Universidad de Granada
- Brousseau G. (1998) *La théorie des situations didactiques*, Recueil de textes de Didactique des mathématiques 1970-1990" présentés par M. Cooper et N. Balacheff, Rosamund Sutherland et Virginia Warfield, La pensée sauvage, Grenoble.
- Brousseau G., Brousseau N., Warfield V. (2002), "An experiment on the teaching of statistics and probabilité", *Journal of Mathematical Behavior*, 20 p. 363-441
- Brousseau G. (2003), "Situations fondamentales et processus génétiques de la statistique", *Cours de la 12ème école d'été de la didactique des mathématiques*, à paraître.
- Chevallard, Y. , (1997), Familière et problématique, la figure du professeur, *Recherches en Didactique des Mathématiques* 17(3) 17-54
- Chevallard, Y. (1985), *La transposition didactique, du savoir savant au savoir enseigné*, La pensée sauvage, Grenoble.
- Chevallard, Y. Wozniak F. (2003), « Enseigner la statistique au secondaire : entre genre prochain et différence spécifique », *Cours de la 12ème école d'été de la didactique des mathématiques*, à paraître.
- Combes M.-C., (1996), *Coordination de L'alternance : enjeux et débats*, Paris, La documentation française.
- Desrosières A., (1993) *La politique des grands nombres – Histoire de la raison statistique*, La Découverte, Paris.
- Franklin, L.A. (1992), "Using Simulation to Study Linear Regression," *The College Mathematics Journal*, 23, 290-295.
- Hesterberg, T. C. (1998), "Simulation and Bootstrapping for Teaching Statistics," *American Statistical Association Proceedings of the Section on Statistical Education*, Alexandria, VA: American Statistical Association, 44-52.
- Kennedy, K., Olinsky, A., and Schumacher, P. (1990), "Using Simulation as an Integrated Teaching Tool in the Mathematics Classroom," *Education*, 111, 275-296.
- Jensen, P. A. (1983), "Microcomputer Forum: Microcomputers in the Instruction of Operations Research", *Interfaces*, 13, 18-24.
- Le Nir M., Oriol JC, (2000), L'enquête : élément central de l'enseignement de statistique en première année GLT in *CNR IUT 2000 Recherche et innovation*, Presses Universitaires d'Orléans, pages 157 à 169.
- Malglaive G., Weber A., (1982), « Théorie et pratique : approche critique de l'alternance en pédagogie », *Revue française de pédagogie*, n°61, pp. 17-27.
- Marasinghe, M. G., Meeker, W. Q., Cook, D., and Shin, T. (1996), "Using Graphics and Simulation to Teach Statistical Concepts," *The American Statistician*, 50, 342-351.
- Mills, J. D., (2002), Using Computer Simulation Methods to Teach Statistics: A Review of the Literature, *Journal of Statistics Education* Volume 10, Number 1
- Oriol J-C, (1999), "L'évaluation en alternance", in *Actes du colloque sur l'alternance*, IUT Lumière, Lyon.
- Oriol J-C, (2002), "Réaliser une enquête par questionnaires : un outil didactique pour la statistique inférentielle à l'Université", in *3ème Rencontre Europe-Amérique-Latine sur la formation technologique et professionnelle*, La Havane, Cuba.
- Oriol J-C, Régnier J-C., « Fonctionnement didactique de la simulation en statistique : Exemple de l'enseignement du concept d'intervalle de confiance », *XXXVèmes Journées de Statistique, Lyon 2-6 Juin 2003*, Tome 2 Pages 743 à 750.
- Pedhazur, E. J. (1997), *Multiple Regression in Behavioral Research: Explanation and Prediction* (3rd ed.), Orlando: Harcourt Brace College.
- Pulley, L. B., and Dolbear, F. T. (1984), "Computer Simulation Exercises for Economics Statistics," *Journal of Economics Education*, 3, 77-87.
- Raynal F., Reunier A., (1997), *Pédagogie : dictionnaire des concepts clés*, Paris, ESF Editeur , 408 p.

- Régnier, JC, (1988), « Étude didactique d'une méthode d'apprentissage fondé sur le tâtonnement expérimental de l'apprenant », *Annales de Didactique et de Sciences Cognitives, séminaire de Didactique des Mathématiques* de ULP Strasbourg , pp 255-279
- Régnier, JC, (2000a), « L'alternance: des évidences (en) aux questions », *Forum, revue de la recherche en travail social*, n°93, pp 38-47 ISSN 0988.6486
- Régnier, JC, (2000b), *Auto-évaluation et autocorrection dans l'enseignement des mathématiques et de la statistique*. Note de synthèse pour l'obtention de l'Habilitation à Diriger des Recherches. Université Marc Bloch (13 décembre 2000) 240 p.
- Revuz A., (1980), *Est-il impossible d'enseigner les mathématiques ?*, Paris Puf.
- Ronan C., (1988), *Histoire mondiale des sciences*, Paris Seuil, 704 p.
- Rousset P., (1997), «L'IUT Lumière : une expérience d'IUT en alternance», communication présentée dans l'atelier formations professionnelles du secteur tertiaire et alternance aux journées d'études sur la professionnalisation dans l'enseignement supérieur.
- Vergnaud, G., (1991) La théorie des champs conceptuels, *Recherches en Didactique des mathématiques*, 10/2.3, Grenoble, La Pensée Sauvage Editeurs pp. 133-169.
- Vergnaud, G., (1994) Le rôle de l'enseignant à la lumière des concepts de schème et de champ conceptuel, in M. Artigues, R. Gras, C. Laborde, P. Tavinot (EDS), *Vingt ans de didactiques des mathématiques en France*, Grenoble, La Pensée Sauvage Editeurs, pp. 177-191.