**Janine ROGALSKI et Marc ROGALSKI**

# CONTRIBUTION À L'ÉTUDE DES MODES DE TRAITEMENT DE LA VALIDITÉ DE L'IMPLICATION PAR DE FUTURS ENSEIGNANTS DE MATHÉMATIQUES

**Abstract.** The paper presents a set of results about how do advanced mathematics students (applying to be mathematics teachers) process with specific implication cases. The aim is to identify, if they exist, the kinds of organisation in the use of logic in theirs evaluating the truth value of implications. Several types of implications were used, for identifying factors involved in the management of the premise. Students' profiles were defined depending on the orientation of their answers to factual not computable implications with (always) false premise: 'Logic' ('*the assertion is true as the hypothesis is false*'), 'Relevance' ('*the assertion is stupid*', '*non sense*'), 'Falseness' ('*hypothesis always false, then assertion false*'), 'ND' (non dominant type of answer). The profiles were related to answers to other types of implications, including computable (mathematical) implications, with premise always false or predicates false for some values ("*hors-sujets*": Legrand, 1990). Two populations were compared (107 students in 1999, 71 in 2001). Globally, the distribution of profiles was similar in the two groups ; answers to most of mathematical items differ depending on the profile. The "*logic*" profile was the only one with stable correct responses: it was a minority. Profiles seem to be relatively stable individual properties, enabling some qualitative anticipation as regarding their answers to a series of tests. Even if advanced mathematics students were more logical than subjects questioned in psychology experiments, the frequency of inefficient profiles ('f*alseness*' and '*ND*') led to discuss how profiles could be changed, through scholar teaching and during teachers training.

**Résumé.** On présente un ensemble de résultats sur une étude concernant le traitement d'implications particulières par des étudiants préparant le concours de recrutement de professeur de mathématiques de l'enseignement secondaire (CAPES). Le but est d'identifier, s'ils existent, des types de structuration de l'usage de la logique dans une de ses fonctions : l'évaluation de la validité d'implications. Des implications de différents types ont été utilisées. On a défini des profils à partir des réponses à un ensemble d'implications non calculables à prémisse toujours fausse et mis en relation ces profils avec les réponses à d'autres catégories d'implications, dont des implications (à contenu mathématique) calculables, à prémisse toujours fausse, ou comportant des "hors-sujet" (Legrand, 1990). On a comparé deux populations (107 étudiants passant le test en 1999, 71 en 2001). Globalement, on observe une distribution analogue des quatre profils dans les deux tests ; des résultats différenciés selon les profils (de la même manière dans les deux populations) sur la plupart des items à contenu mathématique ; un profil "logique" est le seul stable dans les réponses correctes : il est minoritaire. On relève une utilisation significative de la contraposition dans les deux populations en ce qui concerne les items comparables (un étudiant sur deux l'utilise au moins une fois). Les profils que nous avons définis nous paraissent des propriétés des sujets qui permettent certaines anticipations qualitatives pour certaines épreuves. Nous discutons de la manière dont ces propriétés des sujets pourraient évoluer, en particulier à travers les processus d'enseignement universitaire et lors de la formation des maîtres.

**ANNALES de DIDACTIQUE et de SCIENCES COGNITIVES**, volume 9, p. 175 – 203.

Mots clés : Didactique des mathématiques, formation des enseignants, logique, raisonnement, erreur.

## 1. Introduction

Nous présentons un premier ensemble de résultats d'une étude de nature empirique faite sur deux groupes d'étudiants licenciés de mathématiques entrant dans la préparation au Capes de mathématiques à l'Université de Lille, la plus nombreuse en septembre 1999 (test 1: 107 étudiants) et la deuxième en septembre 2001 (test 2 : 71 étudiants).

La motivation de l'étude réside pour nous dans l'importance, pour des enseignants de mathématiques en collèges ou lycées, de savoir distinguer dans les "erreurs de raisonnement" des élèves celles qui proviennent directement de la mauvaise compréhension des mathématiques enseignées, et celles liées à un maniement erroné de la logique en oeuvre en mathématiques ; par exemple, un certain nombre d'élèves et d'étudiants pensent spontanément, en actes, que $P \Rightarrow Q$ et $\text{non}P \Rightarrow \text{non}Q$, "c'est pareil", ou que la négation - souvent d'ailleurs dénommée "contraire" - de $\forall x\ P(x)$ est $\forall x$ non $P(x)$, ou distinguent mal une implication et sa réciproque " dire $P \Rightarrow Q$ et dire $Q \Rightarrow P$ c'est dire la même chose" (Une étude extensive sur de bons élèves du niveau de 4ème - 8th grade - met en particulier ce point en évidence: Küchemann & Hoyles, 2002). L'origine de ces erreurs est souvent plus cachée, et nous faisons l'hypothèse que les enseignants ne seront vraiment aptes à les détecter et à proposer des remédiations éventuelles que si eux-mêmes ont des idées claires sur le sujet (cf Durrand-Guerrier, 1996; Durrand-Guerrier et al., 2000). Ce sont donc ces idées que nous nous proposons d'étudier sur ces futurs enseignants.

Précisons ce que nous entendons par "idées". Nous ne cherchons pas à repérer, comme c'est la dominante dans les études de psychologues sur le raisonnement à partir du travail pionnier de Wason (Wason, 1966), des modèles généraux de processus de raisonnement logique. Nous voulons essentiellement mettre en lumière, s'ils existent, divers types de structuration plus ou moins inconsciente de l'usage de certains aspects de la logique, en en détectant des schèmes d'utilisation. De plus, la qualité de mathématicien des sujets étudiés, et la nature assez mathématique des tests proposés, font que les très nombreux résultats classiques des psychologues sur le caractère rationnel ou pas des humains (Stanovich, 1999 ; Stanovich & West, 2000) nous paraissent peu pertinents pour ce que nous cherchons à voir. Par exemple, et bien qu'il ne s'agisse pas d'une question directement formulée mathématiquement, les taux de succès au test classique de sélection de cartes de Wason (cf annexe) sont incomparablement plus élevés dans les deux populations étudiées que dans toutes les populations sur lesquelles les psychologues ont travaillé, y compris des universitaires scientifiques (Kern, Mirels

& Hinshaw, 1983; Tweney & Yachinin, 1985, Politzer, 2001).

Dans un premier temps, et c'est l'objet du travail présenté ici, nous nous sommes concentrés sur l'implication, et en particulier sur les divers modes de traitement de la validité de l'implication que ces étudiants "avancés" pratiquent.

Nous avons laissé de côté pour l'instant l'étude de "modes de raisonnement mathématiques" plus globaux, que nous reprendrons éventuellement plus tard : nous travaillons plutôt sur "le pas de raisonnement" au sens de R. Duval (1991), mais en prenant en compte aussi bien des pas de raisonnement sans quantificateurs, qu'avec un ou plus d'un quantificateur (voir à ce sujet les travaux de V. Durand-Guerrier et ceux de M. Legrand). Échappent ainsi à notre étude l'ensemble du processus de démonstration, les méthodes de recherche d'une démonstration, et les processus de recherche de conjectures. Un certain nombre de travaux sont consacrés à cet aspect des choses (Balacheff, 1987 ; Boero, 2000 ; Dreyfus, 1997 ; Hanna, 2000 ; Raffali & David, 2002), et il faudra sans doute développer des études sur les liens entre les activités "raisonner", "chercher", "démontrer", d'une part, et l'utilisation de la logique, de l'autre. Nous pensons néanmoins que le "pas implicatif" en mathématiques est une forme d'inférence constamment présente dans tous les raisonnements heuristiques et toutes les argumentations en jeu dans une recherche mathématique, bien avant la mise en forme rigoureuse de la démonstration. Anticipant sur les résultats qui suivent, nous ajoutons qu'il faudra aussi s'engager dans une prise en compte du contexte d'ensemble de ces activités de raisonnement (niveau des sujets -élèves, étudiants, enseignants-, visée du travail mathématique, contenu mathématique et rapports des sujets à ce contenu mathématique, types d'implications en jeu, visée didactique elle-même).

Dans cet article, nous étudions en particulier les modes de traitement par les étudiants des implications à hypothèse fausse. Pourquoi nous intéressons-nous à cet aspect ? Nous avançons plusieurs raisons.

(1) La situation de l'implication à hypothèse fausse est à peu près la seule qui met fortement en évidence le saut qualitatif entre d'une part une certaine logique usuelle, naturelle, en mathématique, où on se contente souvent de déduire une propriété universellement vraie d'une autre, en utilisant une implication dont l'hypothèse est donc, par l'objectif même qui est visé, vraie, et d'autre part une logique formelle dont la nécessité intervient de façon plus cachée en mathématiques, au point que souvent les étudiants ne se rendent pas compte qu'elle y est plus qu'on ne le pense à l'oeuvre.

(2) La deuxième raison est que justement cet aspect formel de la logique est plus présent qu'on ne le pense dans les mathématiques courantes. Donnons-en quelques exemples.

(*) D'abord, la véracité de $P \Rightarrow Q$ quand P est fausse est nécessaire à la cohérence des raisonnements, en particulier dans les raisonnements par contraposition ou par l'absurde, puisque pour montrer que $A \Rightarrow B$, on va être amené à montrer que nonB$\Rightarrow$nonA, alors même, évidemment, que B peut être vraie, donc nonB fausse !

(*) De plus, les modes usuels de recherche d'une preuve utilisent souvent des conditions suffisantes : "pour prouver Q, il suffirait qu'on ait P, car P$\Rightarrow$Q" : alors même qu'on ne sait pas encore que P est vraie (et il se peut qu'elle se révèle fausse !), il faut "avoir confiance" dans cette implication.

(*) Dans certaines preuves par récurrence, il peut être facile de montrer que $P_n \Rightarrow P_{n+1}$, sans savoir si $P_0$ est vraie, ou $P_1$, … et c'est parfois à partir d'un n assez grand que cela va marcher. Là encore, il faut être sûr de la véracité de l'implication, même si $P_n$ est (éventuellement) fausse. C'est d'ailleurs exactement ce qui se passe dans l'un des items proposés dans nos tests.

Voici un exemple qui cumule les trois situations. On veut prouver que "*pour tout* $n \geq 0$ $4^n+1$ *n'est pas divisible par 3*". Voici un raisonnement assez standard qui demande un certain degré de formalisme : "*sinon, il existe un* $k \geq 0$ *tel que 3 divise* $4^k+1$ *; évidemment* $k>0$ *(3 ne divise pas 2 !). Mais alors il est facile de prouver que 3 divise* $4^{k-1}+1$ *; par récurrence "descendante", on retrouve que 3 divise 2, ce qui est absurde*". Ainsi, toutes les implications du raisonnement de la récurrence descendante sont de fait à hypothèse et conclusion fausses ! (Bien sûr, on peut faire une preuve directe par récurrence "montante", mais la preuve de $P_n \Rightarrow P_{n+1}$ se fait naturellement par contraposée, et on utilise encore une implication à hypothèse (et conclusion) fausse ; prouver sans contraposée $P_n \Rightarrow P_{n+1}$ demande un peu d'astuce : écrire $P_n$ sous la forme $4^n+1 = 3k+\varepsilon$, $\varepsilon = 1$ ou 2, multiplier par 4 ...)

(*) Il y a des cas où la seule preuve connue de A consiste à montrer que non A$\Rightarrow$A (c'est le cas d'un pas indispensable dans la preuve - pourtant très élémentaire quant aux connaissances mathématiques en jeu - du théorème de Cantor : il n'y a pas de surjection de X sur P(X) ).

(*) Enfin, il y a des raisonnements, en particulier dans certaines récurrences, qui commencent par : "*Si n=0, il n'y a rien à montrer ; si* $n \geq 1$*...*" qui laissent souvent les étudiants perplexes : que signifie ce "il n'y a rien à montrer" dans une "démonstration" ?

Il ne nous paraît donc pas sans objet d'étudier chez de futurs enseignants le fonctionnement de l'implication à hypothèse fausse. Nous verrons d'ailleurs dans cette étude à quel point un nombre important d'étudiants sont déstabilisés lorsqu'il prennent conscience que dans l'implication universelle "$\forall$ x P(x) $\Rightarrow$Q(x)" il peut y avoir des x pour lesquels P(x) est fausse : voir plus loin la question des "hors-sujet".

Détaillons maintenant les hypothèses qui sous-tendent l'exploitation empirique des données recueillies. La <u>première hypothèse</u> que nous avons faite est qu'on peut avoir accès aux schèmes d'utilisation de l'implication présents dans la population étudiée et aux effets de ces schèmes en dépouillant les réponses des étudiants à des items qui, soit sont des questions de <u>validation d'implications dont</u>

l'hypothèse est toujours fausse, soit attirent l'attention des sujets sur la valeur de la ou des variables qui rendent fausse l'hypothèse - les "hors-sujet" de M. Legrand (1990).

La méthode choisie est la suivante. Nous avons fait passer un test non anonyme aux étudiants en cours d'inscription au Capes au mois de septembre, réunis dans un amphithéâtre, pendant une durée de 4 heures (bien sûr, certains sujets sont restés moins longtemps). Le test n'était pas un test à but sélectif, mais visait à être utile aux enseignants pour savoir "où en étaient les étudiants" sur cette question du raisonnement mathématique : cela était explicitement dit aux étudiants. (Un petit biais existe d'ailleurs : les étudiants auxquels nous avons eu affaire à cette période étaient ceux qui n'avaient plus de seconde session d'examen à passer en septembre pour finir leur licence, donc une population peut-être légèrement meilleure que celle formée de la totalité des étudiants inscrits en octobre à la préparation au Capes.) Les items des deux tests étaient très variés, allant de questions concernant en apparence la vie courante à d'autres très mathématiques (on pourra voir en annexe 2 les items analysés ici).

La deuxième hypothèse faite, pour le dépouillement, est qu'on peut détecter l'éventail des structurations des modes de validation de l'implication par l'étude des conjonctions de réponses à certains types d'implication présents dans plusieurs items. Ceci nous a amené à établir une typologie de nos diverses implications au moyen de catégories prenant explicitement en compte les contenus mathématiques ou non mathématiques, tant des assertions en jeu que des modes de validation possibles. Nous avons ainsi distingué :

- les implications "*calculables*", à des degrés variés (et donc à contenu mathématique) : "$H_n \Rightarrow H_{n+1}$", "si 1=2 ..", "si $(x^2+1) \leq 0$", les deux items à contenu polynomial,
- les implications "*arbitraires*", correspondant à la définition d'une "règle" (en général non mathématique) : "Wason", "Radford",
- les implications de "*contrat social*" : "les bonbons de la maîtresse",
- les implications "*factuelles*", *non calculables*, où les assertions P et Q (hypothèse et conclusion) sont des données immédiatement saisissables par le sujet, pratiquement sensibles, et qui peuvent éventuellement être à contenus mathématiques "routiniers" pour les sujets : "Triangle", "Circuit", "Labyrinthe".

Nous n'avons pas proposé d'implications sans lien de sens entre P et Q (du type : "si 1=2 alors les escargots ont des ailes"?).

## 2. Les résultats au premier test : existence de profils types, corrélations de ces types avec le succès à certains types d'items

Nous avons lors du dépouillement du test proposé à la première population (Rogalski & Rogalski, 2001) classé les sujets selon quatre profils de comportement, avec variantes pour les trois premiers, en croisant les réponses à la validation de trois implications de type factuel (voir annexe 2) à hypothèse fausse (en fait : toujours fausse, dans la mesure où il y a une variable universellement quantifiée, parfois cachée) : le circuit électrique de Marc Legrand (Legrand, 1990), le labyrinthe d'Evapm étudié par V. Durand-Guerrier (1996), et le triangle.

Le choix de ces items comme permettant de tester ces profils repose sur leur aspect factuel ou "matériel", leur immédiate intelligibilité par les sujets, l'absence de référence à des connaissances mathématiques non routinisées (autres qu'immédiates pour les sujets considérés), et la non-calculabilité de l'implication : la proposition "conclusion" ne peut être dérivée de l'hypothèse par un calcul, qui permettrait aux étudiants de dérouler un schème d'enchaînement de calculs simples. Aussi bien le sens des propositions que la valeur de vérité de l'hypothèse et de la conclusion sont aisément évaluables par référence aux propriétés des objets en jeu (pour les sujets adultes cultivés de la population concernée).

Les profils obtenus sont les suivants :

**Logique stable** (10 sujets)
(réponse du type : "*l'implication est vraie" -en général avec l'argument "parce que l'hypothèse est fausse"* - dans les 3 items)

**Logique instable** (9 sujets)
(la même réponse dans 2 items sur 3)

**Pertinent stable** (9 sujets)
(réponse du type : "*l'implication est stupide", "elle n'a pas de sens"*, dans les 3 items)

**Pertinent instable** (14 sujets)
(la même réponse dans 2 items sur 3)

**Non conditionnel stable** (22 sujets)
(réponse du type : "*l'assertion est fausse car l'hypothèse est fausse"*, dans les 3 items)

**Non conditionnel instable** (23 sujets)
(la même réponse dans 2 items sur 3)

**Sans dominante** (20 sujets)
(tous les autres types de réponses ou de distribution de réponse aux 3 items, incluant les non réponses éventuelles).

Le terme "non conditionnel" est choisi pour rendre compte du fait que ces étudiants ne conçoivent pas la vérité de l'implication comme un lien conditionnel entre les valeurs de vérité des propositions.

Bien sûr, la question qui se pose alors est la pertinence de cette définition des profils : sont-ils un moyen de prédiction statistique au succès ou à l'échec aux autres items ? Y a-t-il corrélation avec le succès au Capes ? En ce qui concerne la population du test 1, la réponse à ces deux questions est développée dans (Rogalski & Rogalski, 2001), et c'est chaque fois : oui !

Plus précisément (ainsi que le montrent les tableaux en annexe 1), les pourcentages de succès à l'item à implication calculable "$H_n \Rightarrow H_{n+1}$" est supérieure à 65 % dans les trois premières catégories de sujets (et même à près de 80 % dans les deux catégories "stables"), et nettement plus bas (entre 35 % et 48 %) pour les autres. Les pourcentages de succès aux deux items sur Wason et sa variante Radford sont supérieures à 55 % dans les trois premières catégories, et même à 65 % dans les deux catégories "stables", et inférieurs à 42 % dans les autres. Les pourcentages de succès à l'item portant sur un contrat social (la maîtresse) sont encore plus discriminants : 63 % en moyenne pour les deux catégories "logique", et inférieurs à 30 % dans les autres, et même seulement 10 % pour la catégorie "pertinent stable".

Enfin, une corrélation analogue existe pour l'admissibilité au Capes, en référence aux taux moyens au niveau national et dans l'académie de Lille, ainsi que le montre la figure 1 ci-dessous.

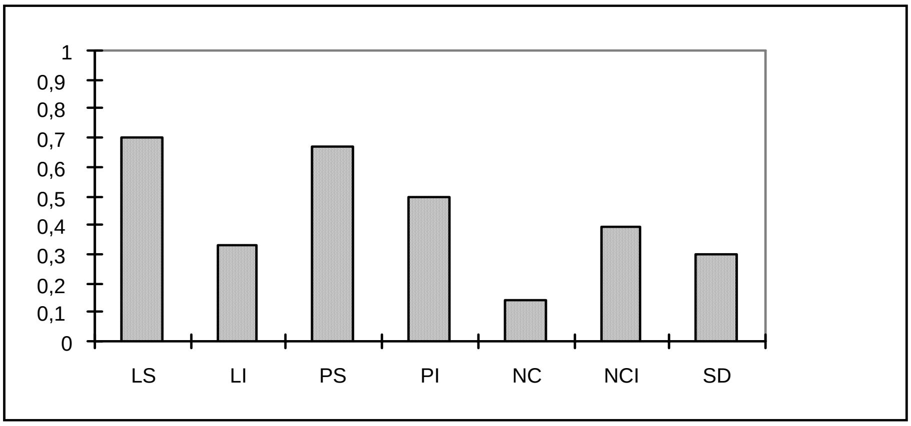



**Figure 1** : Pourcentage des étudiants admissibles en fonction de leur profil.

Au total, 38,3% des étudiants de la population concernée par le test1 sont admissibles (alors que les pourcentages d'admissibles au niveau national et au niveau de l'Académie de Lille se situent autour de 28,5%). Les pourcentages sont nettement plus élevés pour les sujets à profil *logique* ou *pertinent* stable, et nettement plus bas pour le profil *non dominant* ; les autres profils s'éloignent peu ou pas du pourcentage général.

## 3. Confirmation des résultats au deuxième test

Nous nous sommes naturellement demandé si les résultats obtenus avec le premier test étaient relatifs à la population particulière testée, ou s'il y avait stabilité d'une population à une autre (comparable). Nous avons donc recommencé l'opération test en septembre 2001, exactement dans les mêmes conditions, avec un texte reprenant, parmi d'autres, les mêmes items que ceux étudiés dans le premier dépouillement (avec parfois quelques différences de formulation dont nous voulions tester l'effet, en particulier l'introduction de la forme "*si… alors…*", dans certains items). Disons tout de suite que cette modification a été sans effet sur l'item "triangle" ce qui nous a permis de le garder pour la définition des profils, avec les deux autres implications non calculables à prémisse fausse (de formulation inchangée).

Nous avons construit les mêmes catégories avec les mêmes trois items, et aussi étudié la corrélation entre ces catégories et le succès aux mêmes autres items que dans le premier test. La population ayant baissé (moins de candidats au Capes ?), nous avons dû opérer des regroupements pour éviter des dispersions qui auraient rendu les résultats peu significatifs ; précisément, nous avons regroupé les catégories "stable" et "instable" de même type en une seule catégorie ; nous travaillons donc à partir d'ici avec quatre catégories ou quatre profils : "Logique", "Pertinent", "Non conditionnel", et "Sans dominante".

La répartition en catégories de la population testée est alors semblable à celle obtenue avec la première population, ainsi que le montre le tableau 1 ci-dessous.

| **PROFIL** | TEST 1 (N=107) | TEST 2 (N=71) |
|---|---|---|
| LOGIQUE | 17,7 | 21,1 |
| PERTINENT | 21,5 | 23,9 |
| NON CONDITIONNEL | 42 | 39,4 |
| SANS DOMINANTE | 18,7 | 15,5 |

**Tableau 1** : Distribution -en pourcentage-des étudiants selon leur profil de réponse aux implications "non calculables à prémisse fausse".

Remarque : il y a un peu plus de profils "logique" ou "pertinent" dans la population des étudiants du test 2, mais la différence des distributions est très faible.

Les distributions des réponses sur chacun des trois items sont également très proches pour les deux tests, avec la même dominante des profils non conditionnels autour de 40%. Il en est de même pour les distributions et à l'item de contrat social ("les bonbons de la maîtresse")-avec le tiers de réponse de type pertinence où l'étudiant décrit un univers pour lequel l'hypothèse redevient vraie-, et pour l'implication calculable ($H_n \Rightarrow H_{n+1}$) -avec plus de 40% de réponses correctes, et un même taux de confusion avec la démonstration de $H_n$ par récurrence (14%).

De plus, comme le montrent les tableaux 2a, 2b et 2c ci-dessous, les corrélations entre ces catégories et les succès aux mêmes autres items sont elles aussi globalement semblables, à des détails près qui s'expliquent par les variations de formulation, avec cependant quelques différences locales qui soulèvent des questions sur lesquelles nous reviendrons dans la discussion.

| | LOGIQUE | PERTINENCE | NON CONDITIONNEL | SANS DOMINANTE |
|---|---|---|---|---|
| IC: ($H_n \Rightarrow H_{n+1}$) | | | | |
| TEST 1 | 73,7 | 56,5 | 46,7 | 35 |
| TEST 2 | 73,3 | 29,4 | 28,6 | 54,5 |

**Tableau 2a** : Pourcentages de réponses correctes à l'implication calculable à prémisse fausse ($H_n \Rightarrow H_{n+1}$) selon les profils des étudiants, pour chacune des deux populations.

| | LOGIQUE | PERTINENCE | NON CONDITIONNEL | SANS DOMINANTE |
|---|---|---|---|---|
| Contrat social | | | | |
| TEST 1 | 63,2 | 21,7 | 26,7 | 20 |
| TEST 2 | 60 | 35,3 | 35,7 | 18,2 |

**Tableau 2b** : Pourcentages de réponses correctes (selon la logique mathématique) à l'item de contrat social selon les profils des étudiants, pour chacune des deux populations.

| | LOGIQUE | PERTINENCE | NON CONDITIONNEL | SANS DOMINANTE |
|---|---|---|---|---|
| Sélections (Wason & Radford) | | | | |
| TEST 1 | 68,4 | 52,2 | 37,8 | 30 |
| TEST 2 | 66,7 | 52,9 | 60,7 | 36,4 |

**Tableau 2c** : Pourcentages de réponses correctes aux items de sélection (Wason & Radford) selon les profils des étudiants, pour chacune des deux populations.

**Remarques**

- Le *profil logique* se comporte de manière similaire dans les deux tests,
- les réponses correctes à l'item de *contrat social* sont assez voisines pour les deux tests, dans tous les profils,
- on observe une différence marquée dans les réponses correctes à l'*item calculable* ($H_n$=>$H_{n+1}$) pour le *profil non conditionnel*, qui donne nettement plus de réponses correctes dans le test 1 que dans le test 2, alors que la situation est inversée pour les réponses correctes aux deux *items de sélection* (Wason & Radford). Nous verrons plus loin que ce dernier résultat est largement attribuable à l'importance d'utilisation de la contraposée dans l'item Wason dans la formulation classique en "si .. alors",
- pour l'*implication calculable* on observe en fait des différences entre les deux tests pour tous les profils non logiques ; cela ne peut être interprété sans considérer plus précisément la nature des réponses sur l'ensemble de l'item.

Toutefois les différences ne sont guère significatives sur le plan statistique (les Khi2 sur la distribution des réponses donnent $.05<p<.10$, sauf pour la différence pour le profil non conditionnel dans les épreuves de sélection, où $p<.05$). Si on compare les réponses correctes à la fois aux items de sélection et à l'implication calculable, on trouve des résultats qui ne diffèrent pas significativement entre les deux tests, et qui contrastent fortement le profil logique aux autres.

Nous n'avons pas encore à la date de rédaction de ce papier les moyens de comparer avec l'admissiblité et le succès au Capes aux niveaux national et académique, nous ne pouvons faire qu'une comparaison des deux populations (test 1 et test 2) selon les profils (tableau 2d).

| | LOGIQUE | PERTINENCE | NON CONDITIONNEL | SANS DOMINANTE |
|---|---|---|---|---|
| TEST1 | | | | |
| admissibles | 53 | 52 | 35 | 31 |
| reçus | 26 | 30 | 16 | 25 |
| TEST2 | | | | |
| admissibles | 66,7 | 52,9 | 53,6 | 45,5 |
| reçus | 33,3 | 17,6 | 28,6 | 27,3 |

**Tableau 2d** : Pourcentages d'admissibles et de reçus au Capes selon les profils.

Globalement, il n'y a pas une différence marquée entre les deux populations pour l'admission (22,5 %pour le test1 vs 26,8% pour le test 2), alors qu'elle est forte pour l'admissibilité, nettement plus importante dans le test 2 (près de 55%). Le profil "logique" est dans les deux cas le plus performant. Il apparaît en revanche des différences pour les autres profils. Le profil "non conditionnel" a de bien meilleures performances dans le test 2. Quant au profil "pertinent", il présente le plus faible pourcentage de reçus.

Il n'est pas évident de savoir à quoi peut être due la forte atténuation, du test 1 au test 2, de la différence entre les profils pour l'admissibilité au CAPES. Une hypothèse, mais que nous n'avons pas testée, serait que la diffusion pendant l'année 2000 des résultats au test 1 et de leur analyse auprès des enseignants du CAPES a pu avoir un effet sur les enseignants, qui, même inconsciemment, seraient plus intervenus sur les questions de logique : cela pourrait expliquer partiellement une certaine atténuation des différences entre profils entre le passage du test 2 en septembre 2001 et l'écrit du CAPES en mars 2002 ?

## 4. Étude de points plus spécifiques

Les tests comportaient 20 items dans le premier passage, et 18 dans le second. De nombreuses questions pouvaient ainsi être étudiées, concernant par exemple : l'usage des quantificateurs (en particulier en rapport avec la notion de contre-exemple et dans des items utilisant des polynômes du second degré dépendant de plusieurs paramètres) ; les corrélations éventuelles entre les profils et certains items très mathématiques (différence entre suite numérique non majorée et suite tendant vers $+\infty$, par exemple) ; l'effet de certaines formulations "à l'envers" ("nul n'est P s'il n'est Q", "parmi les rationnels, seuls les décimaux peuvent avoir 2 développements décimaux illimités différents" ?). Dans un premier temps, nous avons étudié quatre questions : quelle est l'utilisation de la contraposée ? y a-t-il des effets des changements de formulation ? comment sont traitées les différentes implications calculables à hypothèse toujours fausse ? que se passe-t-il quand on attire l'attention des étudiants sur l'existence de "hors sujet" (au sens de M. Legrand, c'est à dire des x pour lesquels l'hypothèse P(x) est fausse).

### 4.1. L'utilisation de la contraposée pour valider une implication

Sur l'ensemble des 8 items identiques ou à formulation voisine figurant dans les deux tests, nous avons comparé l'usage de la contraposée. Les résultats sont les suivants, en pourcentage des étudiants :

| Utilisent la contraposée | 0 fois | 1 fois | 2 fois | 3 fois | 4 fois |
|---|---|---|---|---|---|
| Test 1 | 59 % | 27 % | 12 % | 2 % | -- |
| Test 2 | 45 % | 24 % | 24 % | 5,6 % | 1,4 % |

**Tableau 3** : Utilisation de la contraposée dans les deux tests (sur les items communs).

Au total, 41 % des sujets utilisent au moins une fois la contraposée lors du premier test, et 55% lors du deuxième test. Le fait qu'un étudiant sur deux, en moyenne, utilise au moins une fois la contraposée est un résultat qui contraste avec une remarque figurant dans (Deloustal-Jorrand, 2000) sur l'absence d'utilisation de la contraposée. Mais son étude clinique concernait un petit nombre d'étudiants, et par ailleurs on peut faire l'hypothèse que les implications à absence de lien sémantique ne déclenchent pas l'usage de la contraposée, même (ou surtout ?) chez des étudiants de mathématiques.

Par ailleurs, il faut noter que la quasi totalité de l'augmentation d'une population à l'autre entre les deux tests est concentrée dans la réponse à l'item "Wason", dont la formulation avait été changée.

| **Types de tâches d'évaluation d'implication** | Test 1 N=107 | Test 2 N=71 | Ensemble N=178 |
|---|---|---|---|
| Implication non calculable (sur 3 items) | 3,1 | 3,7 | 3,3 |
| Implication calculable ($H_n \Rightarrow H_{n+1}$) | 7,5 | 4,2 | 6,2 |
| Vérification de règle : sélection de cartes (Wason) | 7,5 | 29 | ** |
| Vérification de règle : évaluation de procédures | 4,7 | 10 | 6,8 |
| Évaluation de contre-exemples | 23,4 | 32,4 | 27 |

**Tableau 4** : Pourcentage (par item) d'utilisation explicite de la contraposition dans quatre types de situations d'évaluation d'implications dans les deux tests.

** Le fort effet, sur ce seul item, de la différence de formulation ne donne pas de sens à un regroupement des populations des deux tests.

L'évaluation de contre-exemples conduit à une relativement forte utilisation de la contraposition, alors que les autres items d'évaluation d'implication présentent un taux faible (hors Wason avec la formulation "si …alors…"). Si on considère dans chacun des tests le pourcentage d'utilisation de la contraposition dans l'item de sélection de Wason selon le profil des étudiants, on trouve une faible utilisation pour tous les profils dans le test 1 (de 4 à 13%), et une utilisation relativement importante dans le test 2, qui contraste les profils "logique" et "non conditionnel" (de 33% à 40%) et les profils "pertinent" et "sans dominante" (autour de 18%). La mise en relation avec la réponse correcte montre qu'un profil "logique" n'a pas besoin de passer explicitement par la contraposition pour répondre correctement (80% de réponses correctes sans utilisation), qu'un profil "pertinent" peut répondre correctement sans utilisation explicite (60%) et qu'entre 40 et 30% seulement des profils "non conditionnel" et "sans dominante" répondent correctement sans utiliser explicitement la contraposition.

### 4.2. Quelques effets de changement de formulation en "Si ? alors ?"

Trois items ont été modifiés, en introduisant la formulation en "*Si ... alors ...*" :

- "Triangle" version 1 : "tout triangle non aplati du plan dont ... est ..." ; version 2 : "si un triangle a ... alors il est ..."
- "Wason" : version 1 : "derrière une voyelle il y a un chiffre pair" ; version 2 : "si une carte a une voyelle sur une face, alors elle a un nombre pair sur son autre face".
- "Radford" : version 1 : "dans l'urne toutes les boules blanches ont un numéro pair" ; version 2 : "si une boule est blanche alors son numéro est pair".

Les raisons de ces modifications résultent de deux discussions sur la recherche antérieure :

- les travaux de psychologie du raisonnement déductif utilisant l'expression conditionnelle "si ... alors ...", pour comparer notre population de futurs enseignants de mathématiques aux populations de ces travaux nous avons repris leur formulation,
- la formulation avec le terme "tout" ("tout triangle" ou "toutes les boules blanches") pouvaient cantonner des sujets à ne se placer que dans le cas de l'hypothèse vraie, favorisant ainsi a priori un mode d'évaluation de type "pertinent".

Ces changements de formulation ont en fait eu un effet négligeable sur l'item "Triangle" (de 14% à 18% de réponse selon la logique, pas de changement sur les réponses de pertinence ou non conditionnelle) et sur l'item "Radford" (avec davantage de non prise en compte des numéros impairs : de 19% à 30%).

En revanche, le pourcentage des réponses logiques à "Wason" est plus élevé avec la formulation "si …alors …" -il passe de 48% à 67%-, bien que la non prise en compte du nombre impair reste l'erreur dominante, avec plus de 30% des réponses, ce qui accroît d'ailleurs la différence avec les réponses des sujets des expériences des psychologues. De plus, avec l'usage de la formulation conditionnelle "canonique" "*Si ... alors ...*", les distributions de réponses aux deux versions de la tâche de sélection de cartes "Wason" et "Radford" se rapprochent.

Il semble donc que la formulation "canonique" dans Wason ait contribué à déclencher l'usage de la contraposition -à un niveau voisin de celui de l'item de contre-exemples- chez des étudiants qui n'auraient pas répondu correctement sinon : il s'agit particulièrement des étudiants à profil "non conditionnel", comme le montre le tableau 5.

| | LOGIQUE | PERTINENCE | NON CONDITIONNEL | SANS DOMINANTE |
|---|---|---|---|---|
| TEST 1 | 10,5 | 8,7 | 4,4 | 10 |
| TEST 2 | 33,3 | 24 | 32 | 18 |

**Tableau 5** : Utilisation de la contraposition dans la tâche de Wason (implication arbitraire) selon la formulation (test1 : "pour tout ..", test 2 :"si ... alors ...), pour les différents profils.

On observe donc une différence marquée de l'utilisation de la contraposition pour l'item de sélection de Wason pour tous les profils, différence très forte pour le profil "non conditionnel".

Pour vérifier qu'il s'agit bien d'un effet propre au changement de formulation de cet item et non pas d'une différence générale de la population d'étudiants, on a comparé la variation entre le test 1 et le test 2 de l'utilisation de la contraposition dans l'item "Wason" à celle concernant l'ensemble des items communs : on observe bien une augmentation spécifique de l'utilisation pour cet item. De plus, l'augmentation du pourcentage d'étudiants qui utilisent la contraposée pour d'autres items que Wason porte essentiellement sur le profil "non conditionnel". Il semblerait ainsi qu'il y ait une différence (limitée) dans la population des étudiants à profil "non conditionnel" dans le test 2, qui s'ajoute à un effet propre de la formulation, ceci pour tous les profils.

### 4.3. Les différences entre les implications calculables à hypothèse (toujours) fausse

Dans le test 2, trois items sont de ce type : "$H_n \Rightarrow H_{n+1}$", "Si 1=2 alors 2=3", et "Si $(x^2+1) \leq 0$ alors $(x^2+1)^2 \leq 0$". Ils conduisent globalement à un meilleur succès que les implications "factuelles", non calculables, à hypothèse fausse (on observe de 35% à 66% de réponses correctes pour ces items calculables, et de 18% à 34% de réponses correctes pour les items non calculables).

Toutefois ces items diffèrent de manière importante entre eux : par exemple, l'item qui donne lieu au plus grand nombre de réponses correctes est "Si 1=2 ..", alors que le moins bien réussi est "Si $(x^2+1) \leq 0$ ...". En fait, chacun a ses modes de validation (ainsi dans le premier item, la confusion que font certains étudiants entre l'implication $H_n$=>$H_{n+1}$ et la démonstration de $H_n$ par récurrence conduit à un type d'erreur qui n'a pas de correspondant pour les deux autres items).

On peut avancer deux interprétations (partiellement concurrentes) :

- soit il y a une disponibilité de l'invariant qui fait passer de 1=2 à 2=3, qui conduit à un taux élevé de réponses correctes, alors qu'il n'y a pas la même accessibilité de celui utilisé dans l'énoncé pour passer de
$$(x^2+1) \leq 0 \text{ à } (x^2+1)^2 \leq 0,$$

- soit il y a une perturbation introduite par le fait qu'un contre argument au calcul proposé ait été introduit dans l'énoncé de ce dernier item.

Toutes les implications calculables, impliquant un minimum de traitement mathématique, sont mieux "réussies" (évaluation positive) par les étudiants à profil *logique*. Si on prend en compte la cohérence des réponses correctes sur les deux derniers items, la différence est particulièrement forte (60% de réponses correctes pour le profil logique, moins de 30% pour tous les autres), le profil "non conditionnel" ayant le plus faible pourcentage de réponses correctes à ces deux items (10,7%). Nous allons voir que les "hors-sujets" élicitent les mêmes types de différences entre profils.

### 4.4. Les "hors sujet"

Dans la conduite des situations de débat scientifique, M. Legrand a identifié trois types de cas particuliers ou d'événements élémentaires apparaissant dans le débat : des *exemples* qui vérifient l'hypothèse et la conclusion de l'assertion, des *contre-exemples* qui vérifient l'hypothèse et pas la conclusion, et des "*hors-sujet*", qui ne vérifient pas l'hypothèse.

Dans une implication de type "quel que soit x, si P(x) alors Q(x)", il peut y avoir des situations où P(x) est vrai pour certains des x et faux pour d'autres. M. Legrand a proposé d'appeler ces derniers "hors-sujet" pour pouvoir en parler dans les débats, après avoir constaté qu'environ un étudiant sur trois en Deug les assimilait spontanément à des contre-exemples. Cette tendance à traiter ces cas comme des contre-exemples peut conduire ces étudiants à considérer qu'une implication est fausse "quand P(x) n'est pas vraie" (ce qui revient aussi à considérer comme du calcul propositionnel et avec un traitement erroné une affirmation dont la validité devrait être testée dans le calcul des prédicats -cf Durand-Guerrier).

Deux items proposent des implications dans lesquelles existent de tels "hors-sujet" : ils concernent des polynômes du second degré, avec paramètre. Dans un cas (Q5) une question, qui suit la présentation d'un calcul qui permet d'amorcer la validation de l'implication, conduit à identifier un ensemble de valeurs de la variable x pour lequel l'hypothèse est fausse (si l'étudiant y répond sans erreur de technique mathématique). La question suivante porte sur la validité de l'implication. C'est un des items les plus mal réussis dans les implications calculables (environ 20% de réponses correctes), bien qu'il soit d'une forme tout à fait usuelle *a priori* pour un étudiant de mathématique. L'autre item (Q13) qui propose de commenter des réactions d'étudiants est un peu moins mal réussi, mais pose néanmoins problème à une majorité d'étudiants (seulement 41% de réponses correctes).

Si on considère les réponses en fonction des profils, on constate les points suivants :

- les profils "non conditionnel" et "sans dominante" correspondent aux taux les plus bas de réussite, et aucun étudiant à profil "non conditionnel" ou "sans dominante" ne répond correctement à la fois aux deux items de "hors sujet",

- les étudiants à profil "logique" répondent à 87% correctement pour l'item de discussion (n° XIII. 2001), mais un très grand nombre d'entre eux ne répondent pas à la question de l'item V. 2001, montrant l'existence des hors-sujet (conduisant à un faible taux de succès : 33%),
- les étudiants à profil "pertinent' se situent à un niveau intermédiaire.

Alors que les étudiants sont tout à fait susceptibles de "suivre" un calcul montrant que Q(x) est vérifié pour les x vérifiant P(x), ils sont déstabilisés quand on attire leur attention sur l'existence des hors-sujet. Cela nous semble confirmer la fragilité pour nombre d'entre eux du traitement d'une implication en tant que relation entre antécédent et conséquent, fragilité se traduisant par un glissement de l'évaluation de la validité de l'implication à celle de la validité de l'hypothèse (ou de l'antécédent).

## 5. Discussion des résultats

On a analysé deux séries de questions portant sur l'utilisation de la logique du point de vue de l'une des fonctions de la démonstration, rappelées dans (Hanna, 2000), à savoir l'évaluation. Nous avons particulièrement étudié le traitement de la validité de l'implication par des étudiants se préparant pour être enseignants de mathématiques dans l'enseignement secondaire, en accordant une attention privilégiée aux implications à prémisse fausse, que nous avons catégorisées. Un premier test nous a conduit à définir quatre profils à partir d'évaluation d'implications non calculables à prémisse fausse. Lors d'un second test nous avons, d'une part, cherché à évaluer la stabilité des résultats sur une nouvelle population d'étudiants, et d'autre part analysé des points particuliers : utilisation de la contraposition, effets de changements de formulation d'implications, traitement de diverses implications calculables à hypothèse (toujours) fausse, et enfin traitement des "hors sujet".

Globalement, en ce qui concerne la comparaison des populations, on observe une distribution analogue des quatre profils dans les deux tests, des résultats différenciés selon les profils sur la plupart des items et quant à l'admissibilité au concours de recrutement (CAPES).

- Sur l'ensemble des items, les profils "logique" évaluent plutôt correctement les implications (autour des 3/4 des étudiants à dominante "logique"), indépendamment de la valeur de la prémisse - sauf dans le cas où il y a une articulation avec une technique mathématique utilisant une condition suffisante et conduisant à mettre en évidence des "hors-sujet",
- les profils "logique" et "pertinent" apparaissent l'un et l'autre globalement efficaces lorsqu'il s'agit de produire des inférences habituelles (calculer le

passage de P(x) vraie à Q(x) vraie), comme dans les évaluations d'implications quand on se place dans le cas où P(x) est vraie et qu'on ignore le cas où P(x) est fausse, c'est à dire les "hors-sujet". Nos items montrent qu'attirer l'attention sur l'existence de ces "hors-sujet" déstabilise de nombreux étudiants,

- les profils "pertinent" sont beaucoup moins stables dans la donnée de réponses correctes pour les implications calculables : ils sont déstabilisés par la mise en évidence du caractère non vérifié de la prémisse de l'implication, et ils sont assez peu cohérents dans la donnée de réponses correctes à des items voisins. En revanche, ils sont cohérents dans leur réponse à l'item de contrat social ("les bonbons de la maîtresse") où ils construisent le plus souvent un scénario qui rend l'hypothèse vraie, "sauvant" ainsi le contrat posé par la maîtresse,
- les profils "non conditionnel" (l'implication est fausse quand l'hypothèse est (toujours) fausse) sont ceux qui évaluent le plus de manière erronée les implications fausses même calculables (moins de 30% de réponses correctes pour les items pris individuellement, et autour de 10% pour le traitement des hors-sujet),
- les sujets "sans dominante" se situent -en ce qui concerne les réponses correctes-de façon voisine des sujets à profil "non conditionnel" ou à profil "pertinent", selon les items.

Les contrastes les plus forts entre profils se manifestent quand on considère la cohérence de réponses correctes à des items de même classe portant sur un contenu mathématique : implications calculables à hypothèse visiblement fausse, traitement des "hors sujet", cohérence qui est la plus forte pour les profils "logique" et "pertinent", et la plus faible pour "non conditionnel".

Les tâches de sélection issues de Wason ("Wason" et "Radford") différencient moins les profils "logique" et "pertinent" que les tâches dont l'objet est mathématique. Le profil "sans dominante" est celui où on trouve le moins de réponses correctes. Dans l'item de sélection de Wason, l'expression de la règle sous la forme "canonique" : "si ... alors ..." conduit à supprimer la différence observée entre profils "pertinent" et "non conditionnel", en fait elle va de pair avec une augmentation de l'utilisation explicite de la contraposée. Les relations limitées entre les réponses au test longtemps étudié en psychologie cognitive (Wason et ses variantes) et celles d'évaluation des implications utilisées ici conduisent à converger -malgré les différences de populations- avec les conclusions de ceux des psychologues qui, comme Politzer, sont arrivés à la conclusion que la tâche de Wason n'est pas un indicateur significatif de la "rationalité". Les variations et corrélations que nous avons observées indiquent par ailleurs que la question "les humains sont-ils rationnels ou non ?", ou celle "qui est rationnel ?" (Stanovich, 1999) ne tient guère compte de la diversité des modes d'évaluation : la réduction de la question à "rationnel *versus* non rationnel" ne permet pas de rendre compte des différences observées entre les trois profils autres que celui "logique".

Les similitudes globales des deux populations, et les résultats propres au second test sur le traitement des implications calculables et des hors sujets sont une indication de la pertinence des types de profils définis lors du premier test. Toutefois il serait souhaitable de confirmer à nouveau cette validité en déterminant l'appartenance à ces mêmes catégories au moyen d'items différents de ceux utilisés dans nos tests, et d'étudier leur corrélation avec le succès à d'autres items que ceux que nous avons proposés (mais entrant dans les mêmes types d'implications). Les différences observées pourraient être un indice du fait que les conceptions erronées, qui peuvent conduire à des contradictions, conduisent aussi souvent à des réponses peu stables. Reste à expliquer pourquoi les profils "pertinent" et "logique" sont moins souvent corrects dans le test 2 dans le traitement de l'implication calculable liée à une récurrence, alors qu'ils le sont plus souvent pour l'item de contrat social (et ce, sans se différencier entre eux).

L'ensemble des résultats sur les items étudiés montre donc à la fois l'existence de différences quant à des schèmes globaux d'évaluation d'implications, et des effets des types d'implication considérées. La question de l'articulation entre les modes d'évaluation des implications, d'une part, et l'utilisation de pas implicatifs, tant dans les raisonnements heuristiques ou argumentatifs en mathématiques, que dans la construction de démonstrations (c'est à dire la production d'enchaînement d'inférences valides) appelle, nous semble-t-il, la prise en compte de la diversité qui a été identifiée dans les schèmes d'évaluation d'implication (selon les profils) et des sources de variation liées aux types d'implication (dont la typologie devrait être développée et affinée).

La relation entre le traitement de l'implication par des étudiants, candidats à être enseignants de mathématiques, et leur traitement des objets mathématiques a été évaluée ici seulement à travers les réponses correctes à certains items d'implication calculable, et à des "hors-sujets", et à travers les résultats au concours du CAPES. De manière constante, le profil "logique" manifeste une meilleure compétence. Ce fait semble particulièrement marqué dans des items où il faut traiter ce qui peut apparaître comme une contradiction : qu'il s'agisse de l'item "$H_n \Rightarrow H_{n+1}$", où le conflit possible vient du calcul initial et de la démonstration $H_n$ est fausse pour tout n pour un ensemble de valeurs du paramètre, de l'implication calculable où il faut évaluer la réponse évoquée d'un étudiant ou du hors-sujet sur lequel on attire l'attention de l'étudiant. Dans ces cas, le profil pertinent "résiste" mieux que le profil "non conditionnel", tandis que le profil "sans dominante" présente des taux de réussite très hétérogènes.

### 5.1. Perspectives sur la formation des maîtres

La question qui se pose naturellement, puisque sont reçus au CAPES, et donc vont devenir enseignants, des étudiants dont l'aptitude à évaluer des implications pose de sérieux problèmes, est de savoir s'il faut tenir compte de ce fait dans la formation en IUFM, en y assurant une formation à ces questions. L'objectif étant, par exemple, de faire passer dans le profil "logique" les étudiants qui sont dans un autre profil ? Deux questions supplémentaires se posent : quel type de formation, si elle est nécessaire ? et : ne peut-on résoudre le problème avant, pendant la formation universitaire (Deug et licence) ?

En ce qui concerne ce dernier point, nous avons fait un sondage auprès d'enseignants de licence à l'Université de Lille 1. Il en ressort que les enseignants ne parlent quasi jamais de questions de logique mathématique avec leurs étudiants, même "en passant", à l'occasion d'un raisonnement un peu délicat. Il n'y a qu'une exception notable : le raisonnement par récurrence semble effectivement être l'objet d'un certain enseignement, car il est réputé donner lieu à de nombreuses erreurs. Mais, par exemple, aucun des enseignants interrogés ne semble faire de commentaire quand il débute une démonstration par "si $x = 0$, il n'y a rien à démontrer ; supposons $x \neq 0$,?", ils ne s'attendent pas à ce que cela puisse poser un problème aux étudiants. En fait, la logique mathématique que les enseignants du supérieur utilisent leur semble tellement naturelle qu'ils n'imaginent pas que ce ne soit pas le cas pour leurs étudiants.

Ici ou là, on a rétabli parfois en DEUG un enseignement de logique "formelle", mais à la lumière de ce qui se passait-il y a 20 ans, on peut douter qu'un tel enseignement puisse avoir un effet réel sur la pratique mathématique des étudiants. Par contre, deux modes d'action peuvent être utilisés apparemment avec un certain succès.

D'une part, un enseignement ayant pour but une utilité immédiate en mathématique, en particulier permettant l'utilisation du débat scientifique avec des raisonnements validables par une collectivité étudiante, en attirant l'attention sur les points essentiels qui distinguent la logique des mathématiciens de la logique "courante" spontanément pratiquée par les étudiants. C'est par exemple le cas de l'activité "Circuit" développée par M. Legrand à différents niveaux des cursus ; on pourra voir à ce propos (Legrand, 1990).

De l'autre, à travers des options spécifiques de DEUG ou de licence, on peut donner un enseignement "culturel" sur la logique, en plongeant les étudiants, d'une part dans l'évolution historique, la philosophie et l'épistémologie des questions de logique (avec lecture de textes et discussions), de l'autre dans les difficultés proprement mathématiques et didactiques qu'on peut rencontrer chez des enseignants et des étudiants ou des élèves. C'est par exemple ce qu'a pratiqué V. Durand-Guerrier à l'Université de Valence ((Durand-Guerrier, 1996), et ce qu'une expérience récente a mis en oeuvre à l'Université de Nice (Pellissier, 2002).

De toutes façons, on peut s'attendre à ce qu'il soit impossible d'obtenir des universitaires mathématiciens, dans un avenir proche, que la totalité des étudiants titulaires de la licence ait une solide compréhension - pour ce qui nous concerne ici - des conditions de validation de l'implication.

Il semble donc indispensable qu'une mise au point sérieuse sur ces problèmes intervienne en formation des maîtres. Deux difficultés apparaissent. D'une part, les étudiants pensent spontanément qu'arrivés à ce stade, "ils savent raisonner", pourquoi en reparler ? D'autre part, cela leur sera-t-il *utile* , soit pour réussir le CAPES, soit (en deuxième année) pour la conduite de leur classe ? Ces deux difficultés ne concernent d'ailleurs pas que les problèmes de logique ? Nous pensons donc qu'il faut, d'abord, déstabiliser les étudiants d'IUFM quant à leurs certitudes sur la logique. L'utilisation de tests analogues à ceux que nous avons fait passer, avec discussion et mise au point collective, nous semble de ce point de vue une approche possible, voire inévitable, pour rendre les étudiants réceptifs. Il n'est pas sûr, alors, que la seule mise au point à l'issue des discussions suffise à restabiliser les étudiants dans un meilleur "profil". Sans doute les deux points signalés plus haut (différences entre logique courante et logique des mathématiques, et aspects culturels et historiques de la logique) seront-ils nécessaires. Mais il faudra sans doute y joindre une action plus spécifiquement didactique : où, dans quels types de raisonnements, sur quels domaines mathématiques, les élèves de lycée et de collège font-ils des erreurs de logiques, lesquelles, et comment mettre en place des remédiations ?

Autrement dit, une intervention efficace en formation des maîtres sur les questions de logique demandera certainement d'engranger un certain nombre de résultats de recherches à mettre en oeuvre au niveau de l'enseignement secondaire lui-même.

## BIBLIOGRAPHIE

BALACHEFF N., 1987, Processus de preuve et situations de validation, *Educational Studies in Mathematics,18*, 147-176.

BOERO P., 2000, Entrer dans la culture des théorèmes à 12-14 ans : un défi pour la didactique des mathématiques, In T. Assude & B. Grugeon (Éds.), *Actes du séminaire national de didactique des mathématiques*, 41-54, Paris : DIDIREM Université ParisVII.

DELOUSTAL-JORRAND V., 2000, L'implication. Quelques aspects dans les manuels et points de vue d'élèves-professeurs, *Petit x, 55*, 35-70.

DREYFUS, T., 1997, Why Johnny can't prove, *Educational Studies in Mathematics, 28*, 85-109.

DURAND-GUERRIER V., 1996, *Logique et raisonnement mathématique*, Thèse de Doctorat. Université Claude Bernard Lyon 1.

DURAND-GUERRIER V., 1999, L'élève, le professeur et le labyrinthe, *Petit x, 50*, 57-79.

DURAND-GUERRIER V., LE BERRE M., PONTILLE M. C., REYNAUD-FEURLY, J., 2000, *Le statut logique des énoncés dans la classe de mathématiques. Éléments d'analyse pour les enseignants*, Lyon : IREM.

DURAND-GUERRIER V., 1996, Conférence à l'Université de Cornell, USA.

DUVAL R., 1991, Structure du raisonnement déductif et apprentissage de la démonstration, *Educational Studies in Mathematics, 22*, 233-261.

HANNA G., 2000, Proof, explanation and exploration: an overview, *Educational Studies in Mathematics, 44* (1/2), 5-23.

HANNA G., JAHNKE H.E., 1993, Proof and application, *Educational Studies in Mathematics, 24*(4), 421-438.

JOHNSON-LAIRD P.N., Byrne R.M.J, 1991, *Deduction. Hove and London*. UK : Lawrence Erlbaum Associates.

JONES K., GUTIÉRREZ À., MARIOTTI M. A. (Eds), 2000, (Special issue on proof in dynamic geometry environments), *Educational Studies in Mathematics, 44* (1/2).

KERN L. H., MIRELS H. L., HINSHAW V. G., 1983, Scientists' understanding of propositional logic : an experimental investigation, *Social Studies of Science, 13*, 131-146.

KÜCHEMANN D., HOYLES C., 2002, Students' understanding of a logical implication and its converse, In A.D. Cockburn & E. Nardi (Eds.), *PME26.*

*Proceedings of the 26th Annual Conference IGPME*, 3-241-3-248, Norwich: SEPD UEA Norwich.

LEGRAND M., 1989, Genèse et étude sommaire d'une situation co-didactique : le débat scientifique en situation d'enseignement. In C. Laborde (Éd.), *Premier Colloque Franco-Allemand de Didactique des Mathématiques et de l'Informatique*, Grenoble : La Pensée Sauvage.

LEGRAND M., 1990, "Circuit" ou les règles du débat mathématique. In Commission Inter-Irem Université (CI2U) (Eds.), *Enseigner autrement les mathématiques en DeugA première année*, 129-161, ParisVII : IREM.

PELLISSIER R., 2002, *L'enseignement du raisonnement scientifique*, in *Une expérience d'enseignement à Nice*, publication de la Commission Inter-IREM Université (CI2U), à paraître.

POLITZER G., 2001, *Communication Séminaire Cognition & Activités Finalisées*, juin 2001, Université Paris 8.

RADFORD L., 1985, *Interprétation d'énoncés implicatifs et traitements logiques. Contribution à la faisabilité d'un enseignement de la logique au lycée*, Thèse de l'Université Louis Pasteur Strasbourg.

RAFFALLI C., DAVID, R., 2002, Apprentissage du raisonnement assisté par ordinateur, *Gazette des mathématiciens, 92*, 48-56.

ROGALSKI J., ROGALSKI M., 2001, How do graduate mathematics students evaluate assertions with a false premise ? In M. van der Heuvel-Panhuisen (Ed.), *Proceedings of the 25th Conference PME25@NL*, 4.113-4.121, Utrecht, NL : Freudenthal Institute Utrecht University.

STANOVICH K.E., 1999, *Who is rational ? Studies of individul differences in reasoning*, Mahwah, NJ, London : Lawrence Erlbaum Associates.

STANOVICH K.E., WEST R. F., 2000, Individual differences in reasoning : Implications for the rationality debate ? *Behavioral and Brain Science, 23*, 645-726.

TWENEY R.D., YACHININ S.A., 1985, Can scientists rationaly assess conditional inferences ? *Social Studies of Sciences, 15*, 155-173.

WASON P. C., 1966, Reasoning. In B.M. Foss (Ed.), *New horizons in psychology* (Vol.1). Harmondsworth : Penguin.

## Annexe 1 : relation entre profils et réponses correctes aux différents types d'implication

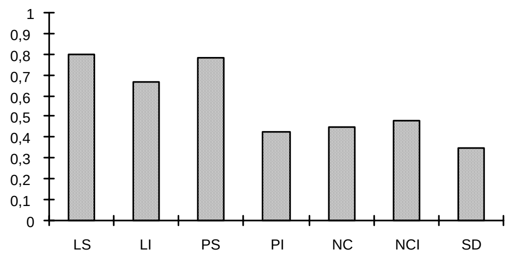



**Figure 2** : Pourcentage d'étudiants donnant une réponse correcte à l'*implication calculable* $H_n \Rightarrow H_{n+1}$, selon leur profil.

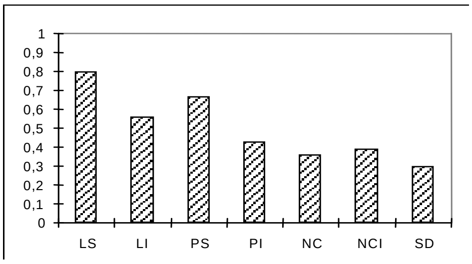



**Figure 3** : Pourcentage d'étudiants donnant une réponse correcte aux *implications arbitraires* (sélection de Wason et Radford), selon leur profil.

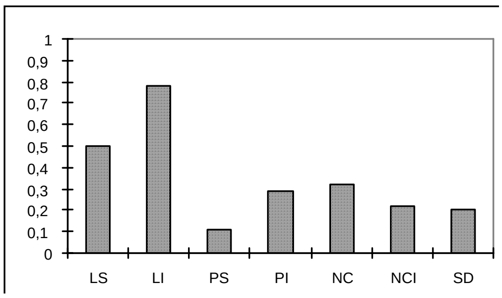



**Figure 4** : Pourcentage d'étudiants donnant une réponse selon la logique à l'*implication de contrat social*, selon leur profil.

**Annexe 2 : Texte des items des deux questionnaires (1999, 2001) analysés ci-dessus**

On donne les numéros des items dans chaque questionnaire, en soulignant ce qui est spécifique de chacun des deux tests.

**Avertissement** (commun aux deux passations) : *les phrases, assertions, énoncés, ... mathématiques qui figurent dans les exercices suivants sont souvent donnés sous forme naïve, non formalisée, comme on le fait dans un texte mathématique courant, voire dans la conversation de tous les jours. C'est volontairement, et vous devez vous sentir très libre dans vos réponses, qui peuvent (doivent !) comporter beaucoup de commentaires, et même si vous le voulez aller jusqu'à : "cette question est idiote !".*

---

## Assertion calculable à premisse fausse

**I. 1999 / 2001**

Etant donné $\lambda \in R$, on considère la suite $(u_n)_n$ définie par :

$$u_0 = \lambda, \quad u_{n+1} = 2u_n + 1.$$

Soit $H_n$ l'assertion " $u_n \leq f(2^n,3) - 1$".

**(a)** L'implication " $H_n \Rightarrow H_{n+1}$" est-elle vraie pour certains n ? pour tout n ?

**(b)** Calculer explicitement $u_n$ en fonction de $\lambda$ et de n [on pourra poser $u_n = v_n - 1$].

**(c)** Montrer que si $\lambda > - f(2,3)$ toutes les assertions $H_n$ sont fausses.

**(d)** Si $\lambda = 10$, que peut-on dire de l'assertion " $\forall n\ H_n \Rightarrow H_{n+1}$ " ?

**(e)** Soit $P_n(\lambda)$ l'assertion "$u_n = 2^{n+5}(\lambda + 1) - 1$". Les assertions suivantes sont-elles vraies ou fausses ?

(1) " $\forall n\ \forall \lambda\ P_n(\lambda)$ est vraie" ;

(2) " $\exists \lambda$ tel que $\forall n\ P_n(\lambda)$ est vraie" ;

(3) " $\forall n\ \forall \lambda\ P_n(\lambda) \Rightarrow P_{n+1}(\lambda)$".

## Contre-exemples

**V. 1999**

Confrontés à l'assertion P suivante :

" Si une fonction f de classe $C^2$ sur R est telle que f et f " soient bornées, alors f ' est bornée ",

deux étudiants de DEUG répondent respectivement :

(1) " L'assertion P est fausse, car la fonction f définie par f(x) = x n'est pas bornée sur R, mais sa dérivée est f ' + 1 qui est bornée " ;

(2) " L'assertion P est fausse, car si $f(x) = x^3$, f et f " ne sont pas bornées, et f ' non plus ".

Que répondrez-vous à l'étudiant (1) ? à l'étudiant (2) ?

### II. 2001

Confrontés à l'assertion P suivante :

" Si une fonction f de classe $C^2$ sur R est telle que f et f " soient bornées, alors f ' est bornée ",

trois étudiants de DEUG répondent respectivement :

(1) " L'assertion P est fausse, car la fonction f définie par f(x) = x n'est pas bornée sur R, mais sa dérivée est f ' + 1 qui est bornée " ;

(2) " L'assertion P est fausse, car si $f(x) = x^3$, f et f " ne sont pas bornées, et f ' non plus ".

(3) "L'assertion est vraie, car si par exemple f(x) = sin x, f et f" sont bornées, et f' est bien aussi bornée".

Que répondrez-vous à l'étudiant (1) ? à l'étudiant (2) ? à l'étudiant (3) ?

## Assertion non calculable à prémisse fausse (triangle)

### III. 1999

Que pensez-vous de la véracité de l'assertion suivante ?

"Tout triangle non aplati du plan, dont les médiatrices ne sont pas concourantes, est équilatéral".

### III. 2001

Que pensez-vous de la véracité de l'assertion suivante ?

"Si un triangle non aplati du plan a ses médiatrices non concourantes, alors il est équilatéral".

## Hors-sujet

### V. 2001

On se propose d'évaluer la véracité de l'assertion (A) suivante :

(A) : "$\forall x \in R\ \forall \lambda \in R$, si $x^2 - 2\lambda x + 2\lambda + 3 \leq 0$, alors $|x| \leq 2|\lambda| + 1$"..

(a) Montrer que l'hypothèse se réécrit sous la forme : $(x - \square)^2 \leq (\square - 1)^2 - 4$.

(b) Que se passe-t-il si $-1 < \square < 3$ ?

(c) Selon vous, l'assertion (A) est-elle vraie ?

## Tâche de sélection de Wason (version classique)

### IX. 1999

On dispose de cartes sur lesquelles figurent sur une face une lettre et sur l'autre un chiffre.

On veut tester la règle éventuelle : "derrière une voyelle il y a un chiffre pair". Pour cela on prend un échantillon de 4 cartes, qu'on pose sur la table ; on voit alors la disposition ci-dessous :

$\boxed{E}$ $\boxed{4}$ $\boxed{K}$ $\boxed{7}$

Quelle(s) carte(s) exactement doit-on retourner pour savoir si la règle est respectée sur l'échantillon ?

### VII. 2001

On dispose de cartes sur lesquelles figurent sur une face une lettre et sur l'autre un nombre.

On veut tester la règle éventuelle : "si une carte a une voyelle sur une face, alors elle a un nombre pair sur son autre face". Pour cela on prend un échantillon de 4 cartes, qu'on pose sur la table ; on voit alors la disposition ci-dessous :

$\boxed{E}$ $\boxed{4}$ $\boxed{K}$ $\boxed{7}$

Quelle(s) carte(s) exactement doit-on retourner pour savoir si la règle est respectée sur l'échantillon?

## Assertion calculable à prémisse fausse (arithmétique)

### XI. 2001

L'assertion "si 1 = 2, alors 2 = 3" est-elle vraie ?

## Version "Radford" de la tâche de sélection de Wason

### XV. 1999

Dans une urne, on dispose d'un certain nombre de boules numérotées. Les boules sont soit blanches soit noires. On s'intéresse à la question suivante : "est-ce que, dans l'urne, toutes les boules blanches ont un numéro pair ?".

On envisage 4 procédures pour répondre à la question :

procédure 1 : on sort de l'urne les boules de numéro pair, puis on regarde leurs couleurs ;

procédure 2 : on sort de l'urne les boules de numéro impair, puis on regarde leurs couleurs ;

procédure 3 : on sort de l'urne les boules blanches, puis on regarde leurs numéros ;

procédure 4 : on sort de l'urne les boules noires, puis on regarde leurs numéros.

Pour chacune des 4 procédures, choisissez parmi les deux options :

(a) la procédure me permettra sûrement de répondre à la question ;

(b) la procédure risque de ne pas me permettre de conclure.

## XII. 2001

Dans une urne, on dispose d'un certain nombre de boules numérotées. Les boules sont soit blanches soit noires. On s'intéresse à la véracité de l'assertion suivante : "dans l'urne, si une boule est blanche, alors son numéro est pair ?".

On envisage 4 procédures pour répondre à la question :

procédure 1 : on sort de l'urne les boules de numéro pair, puis on regarde leurs couleurs ;

procédure 2 : on sort de l'urne les boules de numéro impair, puis on regarde leurs couleurs ;

procédure 3 : on sort de l'urne les boules blanches, puis on regarde leurs numéros ;

procédure 4 : on sort de l'urne les boules noires, puis on regarde leurs numéros.

Pour chacune des 4 procédures, choisissez parmi les deux options :

(a) la procédure me permettra sûrement de répondre à la question ;

(b) la procédure risque de ne pas me permettre de conclure.

## Hors-sujet

### XIII. 2001

On a posé la question suivante à un étudiant : la proposition (P) suivante est-elle vraie ?

(P) : "$\forall(x, y)\in R^2$, si $y^2 - 2xy + 8x + 9 \leq 0$, alors $|y| \leq 2|x| + 4$".

Il a donné la réponse suivante :

"L'assertion (P) est fausse. En effet, l'hypothèse signifie que l'on a $(y - x)^2 \leq (x - 4)^2 - 25$ ; mais si

$-1 < x < 9$, alors $(x - 4)^2 - 25 < 0$, donc l'hypothèse n'est pas vérifiée, et n'impose donc aucune contrainte à y (on est dans la bande $]-1, 9[\mathrm{x}R$) ; mais la conclusion, dans ce cas, entraînerait sur y la contrainte $|y| \leq 22$, ce qui serait absurde".

Que diriez-vous à cet étudiant ?

## Assertion non calculable à prémisse fausse : circuit de Legrand

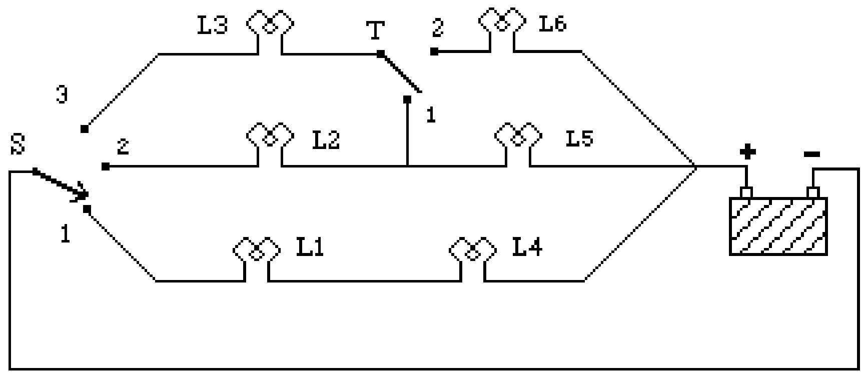
### XVII. 1999 / XV 2001

Le circuit électrique ci-dessous comporte six lampes identiques notées L1, …, L6, et deux commutateurs S et T ; S peut prendre trois positions S1, S2, ou S3, et T peut prendre deux positions T1 ou T2.

(a) Que peut-on dire de la vérité de l'assertion suivante

"si L1 est allumée ou si L6 est allumée, alors L3 est allumée ou L4 est allumée" ?

(b) Que peut-on dire de la vérité de l'assertion suivante "si L1 est allumée et si L3 est allumée, alors L2 est allumée et L5 n'est pas allumée" ?




## Question de contrat social (les bonbons de la maîtresse)
## XVIII. 1999 / XVI 2001

Dans une école primaire, la maitresse donne un jour un problème aux élèves, et leur dit : "cherchez ce problème chez vous ; demain, si quelqu'un a su le résoudre, je vous donnerai des bonbons".

Le lendemain, aucun élève n'a su faire le problème. La maitresse sort un paquet de bonbons et les distribue aux enfants. Ceux-ci protestent : "Ce n'est pas juste, on n'a pas su faire le problème, on n'a pas droit aux bonbons !". Goguenarde, la maitresse leur répond qu'elle a parfaitement respecté le contrat ?

Quels commentaires vous inspire ce récit ?

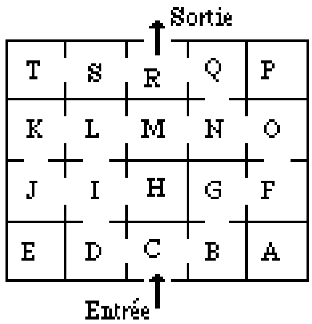
## Assertion non calculable à prémisse fausse : labyrinthe de Durand-Guerrier
## XIX. 1999 / XVII 2001

Une personne nommée X a traversé le labyrinthe ci-dessous, de l'entrée à la sortie, sans jamais être passée deux fois par la même porte. Les pièces séparées par les portes sont notées A, B, …, T.

Pour chacune des sept phrases suivantes, dire si elle est vraie, si elle est fausse ou si on ne peut pas savoir :

(1) X est passé par T ;
(2) X est passé par N ;
(3) X est passé par M ;
(4) si X est passé par O, alors X est passé par F ;
(5) si X est passé par K, alors X est passé par L ;
(6) si X est passé par L, alors X est passé par K ;

(7) si X est passé par S, il est passé par T.




## Assertion calculable à premisse fausse (fonction)

### XVIII. 2001

Un étudiant A affirme que la proposition (I)

(I) : $\forall x \in R$, si $x^2 + 1 \leq 0$, alors $(x^2 + 1)^2 \leq 0$

est vraie, et justifie ainsi son affirmation :

"On a $(x^2 + 1)^2 = x^2(x^2 + 1) + (x^2 + 1)$, donc si $x^2 + 1 \leq 0$, alors $x^2(x^2 + 1) \leq 0$, et $(x^2 + 1)^2$ est la somme de deux quantités négatives, donc est négative".

Un étudiant B conteste son affirmation ainsi :

"Mais c'est absurde ! L'hypothèse $x^2 + 1 \leq 0$ n'est jamais vraie, ni la conclusion $(x^2 + 1)^2 \leq 0$ ! De plus, quel que soit le signe de $x^2 + 1$, son carré est évidemment strictement positif ! Cette assertion est donc complètement fausse !"

Que pensez-vous de ce dialogue ? et de la proposition (I) ?

Janine Rogalski
Laboratoire "Cognition et activités finalisées", Université Paris 8 et CNRS)
Marc Rogalski
(Laboratoire AGAT, Université Lille 1 et CNRS ;
Institut Mathématique de Jussieu, Université Paris 6 et CNRS)
mro@ccr.jussieu.fr