

# Présentation de la base EVAPMIB

## Développement d'une Base de Données d'Évaluation en Mathématiques : EVAPMIB

Antoine BODIN et François COUTURIER  
Institut de Recherche  
sur l'Enseignement des Mathématiques (IREM)  
Université de FRANCHE COMTÉ

Article publié dans le N°5/93 de PEDAGOGIES - Cahiers du laboratoire de Pédagogie Expérimentale de l'Université de LOUVAIN. Numéro spécial consacré à : Évaluation externe et Banques d'Items.

### Pourquoi une base de données d'évaluation

Il y a maintenant près de quinze ans, nous avons mis en place, à l'IREM de BESANÇON, des évaluations importantes au niveau des classes de Collège et de Seconde, créant ainsi un grand nombre de questions d'évaluation expérimentées sur des populations assez importantes. A l'époque, très peu d'évaluations d'ampleur équivalente avaient encore été entreprises en mathématiques dans notre pays (nous ne parlons pas ici de la qualité de ces évaluations, mais plutôt de leur étendue). Nous avons ainsi, à partir de ce moment, commencé à accumuler un certain nombre d'indicateurs concernant les résultats de l'enseignement des mathématiques, au moins dans notre région.

Depuis, nous participons, avons participé, ou sommes associés, à des degrés divers, à de nombreuses évaluations françaises et internationales. En particulier, l'IREM de BESANÇON a pris une part importante dans le développement de l'observatoire EVAPM et est associé de façon indirecte aux évaluations de la DEP ainsi qu'à la Troisième Étude Internationale sur L'Enseignement des Mathématiques et des Sciences (TIMSS de l'IEA).

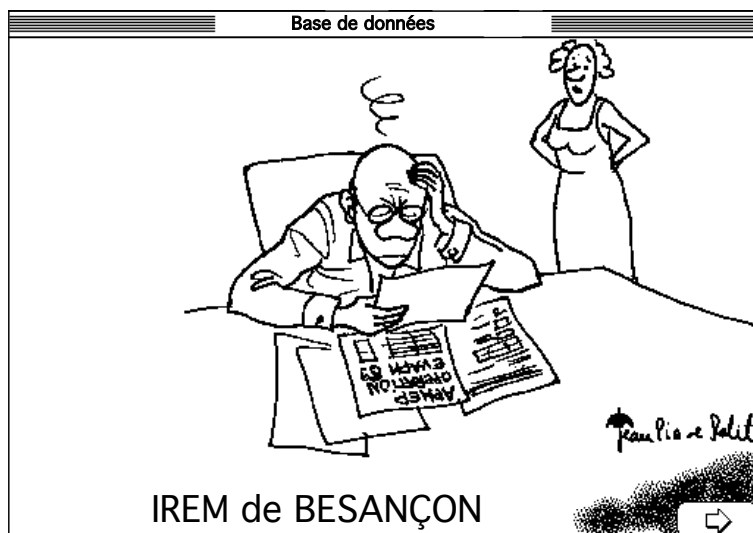
Les évaluations à grande ou moyenne échelle que l'on peut faire sont de peu d'intérêt si elles ne permettent pas de faire des comparaisons dans le temps et dans l'espace géographique.

Face à une question traduisant un objectif de l'enseignement, il y a en fait deux types de temps à prendre en considération :

- 1 - Comment le comportement des élèves d'un même niveau scolaire et d'un même système scolaire s'est-il modifié au cours des  $n$  dernières années ?
- 2 - Comment le comportement des élèves d'une même cohorte se modifie-t-il au long de leur scolarité ?

En ce qui concerne l'espace géographique, il n'est pas indifférent de savoir que tout ce que l'on observe dans un système d'enseignement donné n'est pas universel, mais qu'il y a des constantes et que certains savoirs résistent plus que d'autres aux variations curriculaires.

La fixation sur des taux de réussite ou encore le classement de pays ou régions dans des évaluations ne présente bien sûr pas beaucoup d'intérêt. Ce sont les différences enregistrées, qu'il est tout à fait abusif de rapporter à une échelle unidimensionnelle, qui méritent d'être étudiées et que l'on peut chercher à expliquer. Les évaluations à grande ou moyenne échelle et les informations qu'elles apportent sont sans doute utiles aux responsables de la gestion des systèmes d'enseignement. Elles peuvent aussi, et de plusieurs façons, intéresser la recherche en didactique des



mathématiques. Elles peuvent encore aider les enseignants dans leurs pratiques quotidiennes et jouer un rôle dans la formation des enseignants.

Une évaluation bien conduite suppose que ceux qui l'organisent aient une bonne connaissance d'autres évaluations, préalables, ou simultanées, portant sur le même niveau et sur des niveaux voisins, et qu'ils aient connaissance des interprétations données aux résultats de ces évaluations.

L'expérience prouve que, même dans le cas d'une institution unique qui souhaiterait rester fermée sur elle-même, cette connaissance et les réinvestissements éventuels ne pourraient se faire dans de bonnes conditions en se limitant à des documents papiers. On peut donner de nombreux exemple de cas où la mémoire s'est quasiment perdue, au moins partiellement, entre deux évaluations.

Dans le cas par exemple de l'évaluation EVAPM3/92 du niveau Troisième, aussi bien au moment de l'élaboration de l'évaluation qu'au moment de l'interprétation des résultats, il aurait fallu avoir simultanément à l'esprit (et sur sa table) plus de 2 000 questions d'évaluation. Ces questions se trouvaient de fait dans une base de données composée d'une quinzaine de documents totalisant 3 ou 4 000 pages, l'accès à une question n'étant habituellement possible que si l'on connaît préalablement cette question.

Qu'il s'agisse d'organiser une évaluation au niveau national, régional, local, voire individuel, la possibilité de disposer d'une base de données informatisée, réunissant un grand nombre de questions provenant d'évaluations diverses, serait indubitablement une aide appréciable.

Précisons tout de suite que nous sommes tout à fait conscients que les démarches utilisées pour des évaluations à caractère formatif, voire diagnostique, ne peuvent pas être les mêmes que les démarches utilisées pour des évaluations à caractère sommatif. Il est tout aussi clair que l'on évalue pas de la même façon des élèves en particulier et la qualité d'un programme d'enseignement. Ces réserves faites, il reste vrai que les mêmes questions peuvent être posées dans différents contextes et que la validité de ces questions et les possibilités que l'on se donne d'interpréter les réponses obtenues est dans tous les cas un enjeu essentiel de l'évaluation.

Bien sûr, chacune des questions d'une telle base devrait être accompagnée d'informations relatives aux utilisations antérieures et en particulier aux résultats obtenus et aux observations faites à ces occasions. Elles devraient encore être accompagnées d'études de validité et d'analyses didactiques diverses.

**Exemple de fiche de présentation d'une question**

---

**EVAPMIB - Questions**

*Pour répondre à ces questions suivantes, utilise les langages indiqués sur la figure.*

*Ecris le détail de tes calculs.*

**Calcule le cosinus de l'angle TIR.**

Cos  $\widehat{TIR}$  =

**CALCULE une valeur approchée à un dixième de degré près de l'angle TIR.**

TIR =

EVAPM4/89 - A3:

cosinus avec ou sans explications : R = 49% NR = 23%

valeur approchée de l'angle : R = 36% NR = 37%

D045

D022

D023

4D116

4D117 4D118

Résultats ANALYSE RECHERCHE 22567

---

**Fiche liée présentant les analyses**

---

**EVAPMIB - Analyses**

**ORIGINE** APMEP

**Sens** Outil

**Type d'activité** Calculatoire

**Forme** Fermée

**Temps en min** 3

**Facilité** 40

**Complexité**

**Contenus**

Angle

Cosinus

Triangle rectangle

Trigonométrie

Le fait d'avoir donné les trois côtés d'un triangle rectangle est volontaire et a joué un rôle distracteur.

Il est certain que si les données avaient été contradictoires (par exemple, relation de Pythagore non vérifiée), le comportement des élèves n'aurait pas été changé. A la réflexion, cette façon de formuler des questions en niant l'existence de limitations en ce qui concerne le nombre de degrés de liberté d'une figure nous apparaît tout à fait discutable et à éviter (sauf à faire précéder la question de "est-ce-que la figure représente un objet existant").

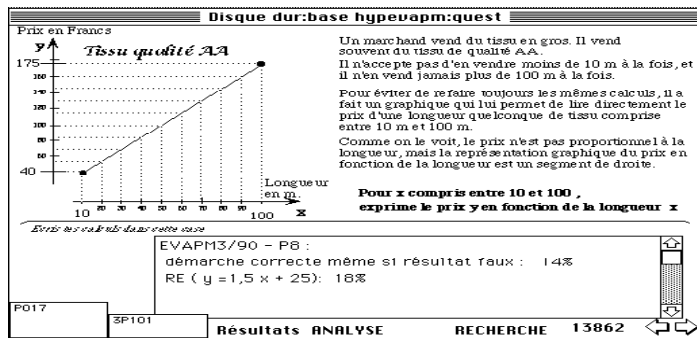
**spécial**

**Recherche**

RETOUR

Le développement de la micro-informatique permet actuellement d'entreprendre ce qui n'aurait été qu'un rêve il y a encore peu de temps : en effet ce développement permet non seulement d'envisager le développement d'une base importante, mais il permet aussi d'envisager de mettre le résultat à la disposition d'un grand nombre d'utilisateurs.

Les responsables d'évaluations institutionnelles montrent souvent une certaine réserve lorsqu'on leur suggère de rendre public les questions d'évaluation qu'ils ont utilisé. Il disent parfois qu'en procédant ainsi "on tue les questions", c'est à dire qu'il ne serait plus possible de les utiliser lors de nouvelles évaluations.



L'expérience prouve qu'après une évaluation, le secret des questions n'est jamais total, et qu'il est bien difficile de connaître la mesure du biais introduit par cette semi-divulgateion. Par contre l'inconvénient du semi-secret est évident : ceux qui auraient besoin de pouvoir analyser les résultats en toute transparence et d'en tirer des enseignements à intégrer à leur pratique (les auteurs de programmes aussi bien que enseignants, les auteurs de manuels, ...) ne peuvent pas le faire de façon

satisfaisante. Pas plus que les médecins (et même les patients) n'accepteraient que les laboratoires médicaux établissent des diagnostics sans leur communiquer la nature des examens pratiqués, pas davantage les enseignants (et même un public plus large) ne peuvent se satisfaire d'analyses globales faites à partir d'évaluations dont ils n'ont pas connaissance. Nous n'aurions pas de difficulté à montrer que certaines thèses, présentées avec d'autant plus d'aplomb que, s'appuyant sur des "instruments de mesure" non divulgués, elles ne peuvent être réfutées, s'appuient en fait sur des instruments ne présentant aucune garantie de validité.

De ce point de vue, la pratique qui s'est instaurée à la Direction de l'Évaluation du Ministère de l'Éducation Nationale (France), de rendre publics, dans la plupart des cas, l'ensemble de documents d'évaluation utilisés nous semble constituer un progrès important.

Ce que nous proposons est de constituer un pool de questions qui puissent servir d'ancres pour de nouvelles évaluations, ces questions étant en nombre suffisamment important pour que le biais introduit par leur caractère public puisse être considéré comme négligeable. Bien sûr, chaque nouvelle évaluation apporterait aussi son lot de nouvelles questions à la fois pour contrôler ce biais et pour élargir la base de données.

De plus, nous souhaitons mettre au point une base de questions aussi diversifiées que possible. En particulier, il n'est pas question de se limiter à des questions à choix multiples ou à des questions fermées dont les qualités psychométriques seraient garanties. Il est maintenant clair que la validité épistémologique et la validité didactique des questions d'évaluation n'est pas correctement prise en compte par les théories classiques de la "mesure" en éducation ; or, sans nous désintéresser d'autres critères possibles, ce sont finalement ces types de validité que nous plaçons au premier plan.

**Question DEP Sixième 1989**

**Base de données EVAPMIB**

Exercice 1

La figure A B C D représente un rectangle.

A. Complète les phrases en utilisant les mots de la liste :  
perpendiculaire, rectangle, parallèle, sommet, isocèle, équilatéral.

a. Les droites (AD) et (BC) sont ..... case 1  
b. Les droites (AB) et (BC) sont ..... case 2  
c. Le triangle AOD est ..... case 3

B. Trouve dans la figure un triangle rectangle.  
Colorie-le. case 4

Evaluation DEP - CE2/6 1989 :

case 1 code1 : 80% (parallèle(s))  
code 3 : 08% (perpendiculaire(s))  
code 9 : 03% (autres réponses)  
code 0 : 09% (absence de réponse)

case 2 code1 : 61% (perpendiculaire(s))  
code 3 : 09% (parallèle(s))  
code 9 : 04% (autres réponses)

**Résultats ANALYSE RECHERCHE 40488**

## Lien avec d'autres bases de données

L'idée de disposer de bases de données pour l'évaluation n'est pas neuve. Dès les années 1970 une importante base était développée aux États-Unis par W.J.POPHAM et son équipe. Depuis, bien d'autres ont dû être créées ici ou là. Citons celles, de langue française, que nous connaissons et qui concernent les mathématiques:

- la banque d'items de mathématiques de l'IRDP de NEUCHATEL (1984),
  - La banque BRUYERE mise au point par Régis GRAS et son équipe de RENNES (voir l'article de Danièle BOINARD dans Petit x n°11/1986),
  - la base d'exercices de l'APMEP qui est consultable sur le serveur APM (36 15 APMEP, pour les non abonnés),
  - la banque de problèmes premier cycle de l'IREM de LYON,
  - la banque d'items de mathématiques du laboratoire de pédagogie expérimentale de l'Université de LOUVAIN (Belgique),
  - la banque d'items de l'I.E.A. (International Association for the Evaluation of Educational Achievement), dont une partie existe en particulier en langue française ;
  - La base de questions d'évaluation propre à l'APMEP (EVAPM). Base n'existant que sur papier.
- ....

Certaines de ces bases sont davantage destinées aux actions de formation qu'à l'évaluation. S'il n'y a pas lieu de confondre base d'évaluation et base destinée à la formation, il n'est cependant pas exclu d'envisager une base d'exercices reliée à EVAPMIB et destinée à proposer des "réponses" aux difficultés rencontrées lors de l'évaluation.

Plusieurs de ces bases utilisent des descripteurs issus de la taxonomie de complexité cognitive et de la typologie d'activités mise au point par R.GRAS, ce qui facilite le passage d'une base à l'autre.

D'une façon générale il serait souhaitable de veiller à la compatibilité des divers descripteurs utilisés, surtout lorsqu'il s'agit de bases à vocation complémentaires.

Le travail de réflexion que nous avons commencé avec une équipe de l'université de LOUVAIN et une équipe de l'IRDP de Neuchâtel va dans ce sens.

Parmi les bases existantes, citons encore MICROCAT, logiciel de "testing" récemment développé aux Etats-Unis par "Assessment Systems Corporation". Mais dans ce cas, la partie recherche et conservation des questions n'est que l'un des modules de l'ensemble ; la mise en relation des informations concernant les questions n'est pas assurée, et la partie gestion générale des situations d'évaluation est largement privilégiée par rapport à la partie proprement didactique.

### Bases informatisées en cours de réalisation

Le "groupe serveur" de l'APMEP développe actuellement une nouvelle base d'exercices pour les premiers et second cycles de l'enseignement secondaire. Cette base, orientée vers la formation, aura des liens avec la base EVAPMIB (en particulier par l'intermédiaire d'EVAPM) ; certaines questions et certains descripteurs seront communs. Les objectifs poursuivis sont à la fois différents et complémentaires.

The screenshot displays a software interface for the EVAPMIB database. At the top, it reads "Question EVAPM Cinquième 1988" and "Base de données EVAPMIB". The main content area shows a drawing of a house labeled "dessin n°1" on a grid. Below it, text states: "Le dessin N°1 a été reproduit à deux échelles différentes, sur le même quadrillage." Below this, a prompt asks: "ECRIS ces ECHELLES dans les cadres correspondants." There are two input boxes labeled "échelle:" with arrows pointing to the drawing. At the bottom right, statistics are shown: "EVAPM5/88 (modalité B) : petit R = 47% N.R. = 20% grand R = 39% N.R. = 21%". The bottom status bar includes "P009", "5F631", "Résultats ANALYSE", "RECHERCHE 8957", and navigation icons.

De son côté, le CNDP a entrepris de développer une base d'exercices de mathématiques pour l'enseignement secondaire. Cette base devrait avoir des éléments communs avec la base APMEP.

## L'état du projet EVAPMIB

A l'IREM de BESANÇON, depuis trois ans, nous travaillons à la réalisation de cette base et une **maquette déjà bien avancée tourne** d'ores et déjà avec plus de 500 questions. Les images illustrant le début de cet article sont des saisies d'écrans provenant de cette base.

Bien sûr, ce travail se fait en étroite liaison avec l'observatoire EVAPM. Une grande partie des questions alimentant la base sont des questions EVAPM, mais l'objectif de la base est aussi, rappelons-le, de nourrir de futures évaluations, EVAPM ou autres.

Le matériel utilisé est du matériel micro informatique Mac INTOSH et le logiciel est HYPERCARD. L'accès à ces matériels est facile dans les universités et particulièrement les IREM, et cela devient de plus en plus souvent possible en d'autres lieux (établissements scolaires et particuliers). De toutes façons, le passage d'un matériel à l'autre n'est plus qu'une question de manipulation technique, et nous nous préoccupons du passage, à terme, sur matériel P.C.

Il convient pour l'instant, du seul point de vue informatique, d'éprouver et d'améliorer cette maquette en concertation avec d'autres équipes et en particulier avec les équipes de recherche de LOUVAIN et de NEUCHATEL et de produire les documents d'accompagnement nécessaires à son utilisation. Simultanément, il nous faut préparer le choix des questions à saisir dans la base ainsi que leur analyse didactique.

La plupart de ceux qui s'intéressent à l'évaluation en mathématiques sont conscient de la nécessité de disposer d'une telle base, mais il certain que seule une base d'une taille importante et connue des chercheurs et des praticiens pourra rendre les services attendus. Le projet est ambitieux et nous savons bien que nous n'en viendront pas à bout tout seuls. et des collaborations multiples sont envisagées. Nous sommes persuadés, qu'à terme, le bénéfice peut être grand pour l'enseignement des mathématiques et le développement de la formation scientifique en général.

D'une façon concrète, nous prévoyons, pour les deux années à venir, des séminaires mensuels de trois demi-journées pour l'équipe de recherche ainsi que trois séminaires de travail en commun avec les équipes Belges et Suisses (une à BESANÇON, une à LOUVAIN et une à NEUCHATEL). Nous prévoyons aussi la consultation et l'intervention systématique de spécialistes de la didactique des divers sujets étudiés, pour assurer le contrôle des analyses didactiques que nous proposerons.

## Analyse d'une question d'évaluation

**Chaque question de la base est accompagnée d'un ensemble d'informations :**

- 1 - Un ensemble de descripteurs.**
- 2 - Des résultats enregistrés lors de passations diverses.**
- 3 - Des valeurs prises par des indices divers.**
- 4 - Des analyses à intention didactique.**

### **1 - Descripteurs et informations diverses**

#### **Numéro de la question**

Il s'agit du numéro de la question dans la base. C'est le seul codage biunivoque : à un numéro correspond une question unique ; à une question correspond un code unique.

#### **Origine et auteur éventuel (= copyright)**

#### **Codes - questions - passation (multi-codage)**

Par définition, et sauf exception précisée et justifiée, les questions présentes dans la base proviennent d'évaluations à grande ou moyenne échelle identifiées et ayant produit des documents de présentation et d'analyse des résultats. Bien sûr, une même question peut avoir été utilisée dans plusieurs évaluations.

Les codes - questions - passation permettant de retrouver la trace des questions dans ces documents originaux : cahiers d'évaluations, rapports d'évaluation....

Dans chaque cas, le code est attribué de façon à mettre en évidence l'évaluation dans laquelle la question a été utilisée et l'identification de la question dans cette évaluation.

*Exemple : Une question posée dans le cadre de l'évaluation EVAPM5/88, Question 17-19 de l'épreuve A, sera codée : EVAPM5/88/A17*

*Cette même question pourra aussi être codée SPRESE3/84/B12 si elle a été utilisée en 1984, dans l'évaluation fin de Troisième organisée par le SPRESE , avec le numéro B12.*

Une même question ayant été utilisée dans plusieurs évaluations sera donc munie de plusieurs codes questions.

### Codes capacités (multi-codage)

Chaque question peut être munie de deux types de codes relatifs aux capacités qu'elle est susceptible d'opérationnaliser :

#### - Les codes capacités des évaluations EVAPM

C'est le cas lorsque la question a été utilisée dans une évaluation EVAPM ou crée dans ce cadre. Ces codes permettent d'utiliser simultanément la base EVAPMIB et les brochures EVAPM.

Soulignons que les formulations EVAPM des capacités correspondaient, au moment où elle ont été introduites, à des formulations officielles des programmes français.

#### - Les codes EVAPMIB.

Pour faciliter un élargissement de la base, ne pas rester liés à des programmes destinés à évoluer, ne pas se limiter à des capacités dites "exigibles", et pouvoir intégrer des questions d'origines diverses, il a été nécessaire de mettre au point un nouveau système de classement des capacités.

Ce système de classement est organisé suivant les contenus en jeu (par thèmes) et est extensible. Pour une part, il intègre, avec des codes différents, les capacités EVAPM.

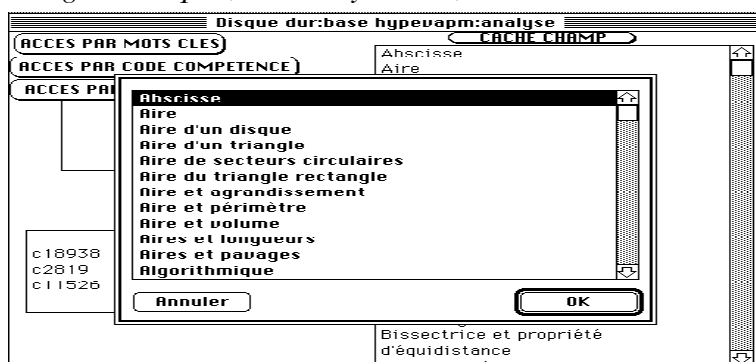
Les formulations EVAPMIB des capacités ne cherchent pas à décrire les questions de façon univoque (elle ne satisfont pas, en général, aux critères "Magériens" de la définition des objectifs) . Ces formulations doivent simplement être considérées comme des descripteurs des questions parmi d'autres.

### Contenus (mots-clé)

A chaque question est associé un certain nombre de mots-clés précisant les contenus en jeu. Ces mots-clés peuvent être des mots isolés : "*Abscisse ; Agrandissement ; Angle ; Arrondi ; Cercle ; Critique ; Décimaux ; Distributivité ; ...*" ou des expressions : "*Calcul numérique ; Construction géométrique ; Axe de symétrie ; ...*".

La liste des mots-clés est extensible en fonction des besoins, mais nous avons évité les synonymes et les expressions dépassant trois mots.

Au niveau de la recherche, les fonctions "ET", "OU", "CONTIENT", offrent la possibilité de former des clés de recherche aussi complexes que l'on voudra.



## Complexité cognitive (au sens de R. GRAS)

Cette taxonomie, qui présente quelque analogie avec la taxonomie de BLOOM et avec d'autres qui ont suivi, a cependant l'intérêt d'avoir été élaborée pour les mathématiques, et donc, de tenir compte des spécificités de la discipline (tableau présentable à l'écran sur demande).

	Objectifs	Activités attendues
	<b>Connaissance des outils de préhension de l'objet et du fait mathématique</b>	
A1	Connaissance de la terminologie et du fait spécifique	Connaître Assembler
A2	Capacité à agir sur une forme physique du concept ou à évoquer	Bricoler, explorer, observer
H3	Capacité à lire des cartes, des tableaux, des graphiques	Dechiffrer, décrire
A4	Effectuation d'algorithmes simples	Calculer, opérer
B	Analyse des faits et transposition	
B1	Substitution d'une démarche représentative à une manipulation Anticipation graphique	Prolonger, induire

## Type d'activité (Typologie R. GRAS étendue)

Les types d'activités attendues des élèves constitue aussi un ensemble de descripteurs qu'il nous a semblé intéressant d'utiliser. La liste de base est celle proposée par R. GRAS, mais cette liste est susceptible d'extension.

Inductive
Heuristique
Classificatoire
Prédicatif
Calculatoire
Logique
Technique
Transfert
Transductif

## Sens (Typologie A. BODIN)

Très souvent les questions d'évaluation portent strictement sur l'objet mathématique sans conduire d'une façon ou d'une autre à le faire fonctionner comme outil (au sens de R. DOUADY). La typologie proposée, très simple, permet de mieux situer la place que prend la question et son éventuelle maîtrise dans le fonctionnement du savoir.

## Forme de la question

Les questions sont aussi classées selon leur forme : fermée, semi-ouverte (QROC), question à choix multiples (QCM), question à réponse rédigée (QRR), question à support matériel autre que papier-crayon-calculatrice (performance item - QSM),...

Fermée
Ouverte
Semi-ouverte
Q.C.M.

## Temps de passation

Le temps proposé est une estimation moyenne et concerne le premier niveau où cette question pouvait être considérée comme "exigible" dans l'enseignement français, au moment où la question a été créée.

## Liste de questions liées

Ce sont des questions qui, présentes dans la base, sont citées dans les analyses, ou qui, présentent un voisinage de contenu important avec la question présentée. Dans une évaluation, certaines de ces questions pourront être considérées comme des alternatives possibles à la question étudiée.

Les listes et tableaux de la plupart des descripteurs présentés ci-dessus peuvent, sur demande, être consultés à l'écran.

Il est possible de faire effectuer une recherche de questions à partir de n'importe quel descripteur ou composition de descripteurs. Bien entendu, il n'est pas nécessaire de s'intéresser simultanément à tous les descripteurs, mais les fiches présentées en consultation renseignent automatiquement sur tous les descripteurs.

## 2 - Résultats enregistrés lors de passations diverses.

Il s'agit de résultats enregistrés lors de passation effectuées dans des évaluations connues, répertoriées (annexe présentable à l'écran), et de représentativités jugées suffisantes...

Les résultats sont constitués des divers pourcentages de réussite ou valeurs "VRAI" recueillies sur un item (voir ce mot). Ils comportent aussi chaque fois qu'ils sont connus, les taux de non-réponses à la question, ou à l'item lorsque cela a un sens (N.R.).

Ces résultats sont présentés dans une fenêtre qu'il est possible de cacher.

Ces taux constituent des indicateurs qu'il convient d'utiliser avec prudence. On trouvera dans les brochures EVAPM et dans les documents papier qui accompagneront la base, des réflexions et des alertes concernant ce point.

### 3 - Valeurs prises par des indices divers.

Il s'agit des valeurs prises par la question pour des indices standards définis sur l'ensemble de la base (d'où des problèmes à résoudre pour des questions ne provenant pas nécessairement des mêmes évaluations).

#### Indice de facilité

L'indice utilisé est gradué de 10 en 10 sur une échelle allant de 0 à 100. Son estimation est basée sur les résultats des passations effectuées au niveau scolaire correspondant au premier niveau où cette question peut être considérée comme "exigible" dans l'actuel enseignement français.

Un indice de facilité de  $n$  signifie qu'au niveau considéré on peut s'attendre à ce que  $n\%$  des élèves réussissent la question (ou du moins une partie importante de la question si cette question est composée de plusieurs sous-questions).

La "Théorie des Réponse aux Items" (Item Response Theorie : IRT) permet le calcul d'indices de facilité (ou de difficulté) indépendants de la population ayant répondu aux questions. Des travaux sont en cours pour étudier l'intérêt et la possibilité pratique d'introduire ce type d'indices dans la base.

#### Indice de discrimination

Le premier indice retenu est du type "Upper-Lower Index". Son calcul pour une question suppose que la question a été intégrée dans une évaluation et que l'on possède, pour cette évaluation, les résultats d'un grand nombre d'élèves. Dans ce cas, la valeur de l'indice de discrimination est égale à la différence entre le pourcentage de réussite à la question obtenu par les 25% des élèves qui réussissent le mieux l'ensemble de l'évaluation et le pourcentage de réussite à la question obtenu par les 25% des élèves qui réussissent le moins bien l'ensemble de l'évaluation.

Les valeurs possibles de l'indice appartiennent donc à l'intervalle  $[-1 ; +1]$ .

Pour une question donnée, la valeur de l'indice de discrimination dépend donc de l'évaluation. Les analyses préciseront si l'indice est stable sur les évaluations connues ; éventuellement plusieurs valeurs de l'indice pourront être données.

Ultérieurement, d'autres indices de discrimination plus indépendants des évaluations pourront être proposés, sans doute basés sur la théorie de réponse aux items (IRT).

#### Indice de généralisation

A partir de la réussite obtenue par un élève à une question donnée, cet indice devra renseigner sur ce qu'il est possible d'inférer (de façon statistique) sur le degré de maîtrise de cet élève par rapport au champ conceptuel concerné par la question (ou du moins une partie identifiée de ce champ).

L'indice en cours d'étude sera basé sur l'analyse implicative développée par R. GRAS,

D'autres types d'indices sont à l'étude tels que : **indice de sensibilité (à l'apprentissage, aux variations curriculaires), indice de lisibilité...**

Ces indices pourront n'être utilisés que sur une partie de la base. De même des courbes caractéristiques de réponse aux items (C.C.I.) pourront accompagner certaines questions.

Comme pour les descripteurs, les indices n'ont pas à être utilisés simultanément ; ils constituent des indicateurs de validité, de fiabilité, ou simplement d'intérêt de la question (nous préférons garder le mot "pertinence" pour les objectifs).

La double vocation de la base, instrument de travail et instrument de recherche, justifie la présence simultanée de descripteurs et d'indices divers dont certains, très simples, ont une signification claire et sont faciles à utiliser, tandis que d'autres, plus sophistiqués sont encore à l'état de prototypes.

### 4 - Des analyses à intention didactique.



Ces analyses peuvent être plus ou moins développées suivant les questions et ne sont pas nécessairement présentes pour toutes les questions.

Pour une part importante, ces analyses sont le fruit du travail de l'équipe EVAPMIB. Ces analyses peuvent cependant être partiellement ou totalement inspirées des analyses présentées dans les documents EVAPM, SPRESE, ou autres. Les parties simplement reprises dans ce type de document seront signalées. Des points de vue différents ou des avis d'experts peuvent aussi être présentés dans ces analyses ; dans ce cas, les auteurs seront nommés.

Une analyse complète peut comprendre :

- une présentation de la question et de son intérêt éventuel,
- des précisions sur ce qui est attendu des élèves assorties d'une analyse sommaire de la tâche (analyse a priori),
- des précisions sur ce qui a été observé au fil des passations,
- un inventaire des stratégies et procédures rencontrées,
- un inventaire des erreurs rencontrées,
- une tentative d'explication des procédures et des erreurs,
- un examen de la validité de la question,
- une liste de questions liées, et des indications sur les corrélations existant entre ces questions et la question présentée,
- une série d'interrogations soulevées par l'étude de la question.

L'analyse d'une question n'est strictement liée ni à une évaluation particulière (EVAPM3/90..SPRESE4/84,...), ni à un niveau scolaire particulier. Cette analyse n'évite ni la critique de la formulation de la question, ni la critique des objectifs qu'elle est sensée opérationnaliser, ni l'implantation de ces objectifs à tel ou tel niveau du curriculum.

Nous essayons d'éviter de figer l'analyse en laissant croire qu'elle est intrinsèque à la question alors qu'elle dépend des informations que l'on possède. Dans une base de données de questions destinées à être réutilisées, les analyses sont nécessairement évolutives.

Les pourcentages de réussite parlent parfois d'eux mêmes. Des commentaires ne sont ajoutés que dans la mesure où ils nous paraissent apporter des précisions ou des nuances utiles.

## Exemple d'analyse de question

### Exemple 1

Question opérationnalisant une capacité dite "exigible" du programme de Cinquième 1988: "compléter un graphique représentant une situation de proportionnalité, les données étant partiellement fournies".

La question a été posée à plusieurs niveaux et à plusieurs reprises et la capacité en question se trouve opérationnalisée dans la base de plusieurs façons.

#### Analyse proposée

Au niveau Cinquième (EVAPM5/88), peu d'élèves ont tracé la demi-droite pour placer les croix. Il semble que la plupart des élèves qui ont réussi cet exercice l'aient fait par le calcul (recherche du prix de 1 kg) ; l'usage de la calculatrice ayant favorisé cette stratégie. A ce niveau, la question a donc une validité assez limitée.

Au niveau Quatrième (EVAPM4/89), on observe une amélioration des taux de réussite mais aussi un changement dans les procédures utilisées. Le passage par le calcul devient rare. Les élèves tracent la demi-droite et placent ensuite les croix demandées. Cette évolution est significative d'une acquisition (théorème en acte ou théorème explicite ?) : "les points de la représentation graphique d'une

relation de proportionnalité sont alignés entre eux et alignés avec l'origine des axes". S'il s'agit de savoir si les élèves ont une bonne maîtrise de cette représentation, il est clair que la question présente une validité plus grande au niveau Quatrième qu'au niveau Cinquième.

En Cinquième comme en Quatrième, la réussite est meilleure dans le sens (objet ----> image) que dans le sens (image -----> objet), cela est assez bien connu, et cela est vrai aussi bien dans le cas du passage par le calcul que dans le cas de la lecture sur une représentation graphique. On trouve la trace de cette difficulté plus loin dans la scolarité (difficulté à utiliser une représentation graphique pour des résolutions approchées d'équations ou d'inéquations).

Compte tenu de la difficulté que semble présenter cette question pour les élèves de Cinquième (1988), la capacité opérationnalisée mériterait d'être accompagnée, et précédée dans les programmes, de capacités telles que : "réaliser un graphique représentant une situation de proportionnalité" et/ou "lire un graphique représentant...".

### Question N° 98389

Code Base EVAPMIB : P018

Capacité code EVAPM : 5P625

Codes question : EVAPM5/88/A25-27 ;  
EVAPM4/89/D14-16

Codes questions liées : EVAPM5/90/N10-12 (autre formulation),  
EVAPM4/89/P28 ; SPRESE3/90/A2  
repris EVAPM3/92/T6)

Origine : APMEP

Mots clés contenus : *Proportionnalité - graphique*

Sens : outil

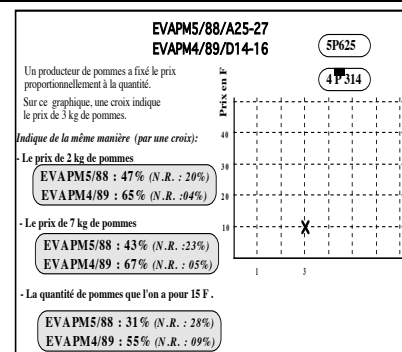
Type d'activité :

Forme : semi-ouverte

Temps : de 2 à 5 minutes

Facilité : 10

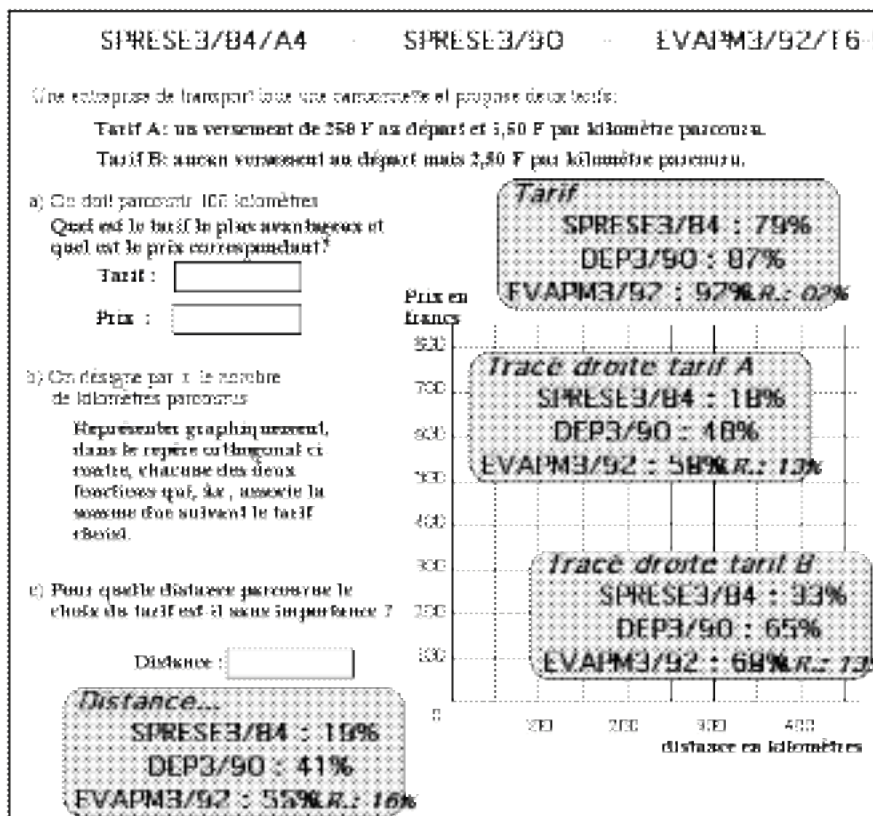
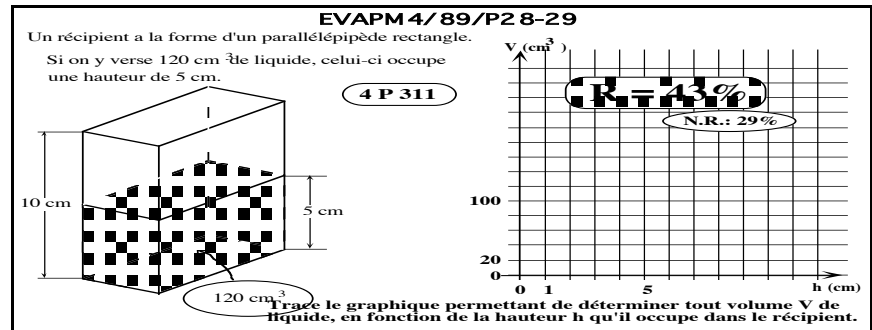
Complexité : B3





La question a été reprise dans EVAPM5/90 sous une forme modifiée (EVAPM5/90/N10). Dans cette opérationnalisation les informations données sont largement redondantes et l'élève est largement aidé par la consigne préalable de "tracer le graphique". L'interprétation est faite ensuite avec peu de difficultés, mais on notera que la difficulté de lecture dans le sens (image -----> objet) subsiste largement.

D'autres questions de la base opérationnalisent la même capacité ou des capacités voisines : EVAPM4/89/P28-29 - SPRESE3/90/A2 (repris EVAPM3/92T6-9) - Ces questions permettent d'avoir une idée de l'évolution de cette capacité dans le temps.



## Quelques définitions

### Question d'évaluation

Une question est une *unité d'évaluation* indépendante, se suffisant à elle-même.

**A propos des notions de question et de réponse.**

Les questions d'évaluation, que nous nommerons simplement **questions** sont les unités de la base de données d'évaluation.

Par habitude, nous gardons la dénomination "question", mais en fait les unités d'évaluation peuvent ne pas conduire à des questions formelles.

Par exemple "Trace deux droites parallèles" est une formulation qui ne contient pas, formellement, de question. Ce ne serait pas le cas avec "peux-tu tracer deux droites parallèles ?", mais dans ce cas, la **réponse** "oui" serait une bonne réponse. Dans les deux cas on pourrait parler de demande ou de stimulus avec comportement réponse attendu.

Mais on peut facilement imaginer des unités d'évaluation sans comportement réponse attendue. La situation d'évaluation peut ainsi consister en la mise en place d'une activité (projet, jeu, coopération entre pairs,...) sans produit privilégié, mais permettant l'observation des comportements. Bien sûr, la situation ne peut être acceptée comme situation d'évaluation que si elle autorise des inférences sur les compétences.

Le mot réponse lui même est conservé par habitude, mais il est ambigu. Il faudrait parler tantôt de produit, tantôt de comportement. Une "question" telle que "trouver une fonction vérifiant telle ou telle propriété", peut n'intéresser l'évaluation que pour ce qu'elle peut révéler des conceptions de l'élève relativement au concept de fonction. Ce n'est pas le produit qui importera, mais bien les procédures mises en oeuvre.

### **Sous-questions**

Certaines questions comportent des sous-questions plus ou moins enchaînées.

Nous évitons d'utiliser les mots "exercice", "problème",... Les questions d'évaluation placées dans la base ne sont pas, a priori, destinées à exercer les élèves. Certaines questions peuvent être des problèmes au sens strict, mais ce n'est que rarement le cas.

### **Item**

L'item est l'unité d'information élémentaire. Un item est une fonction à trois valeurs : "VRAI", "FAUX" ou "Non Concerné".

Une question d'évaluation peut comporter un ou plusieurs items.

Si l'évaluation devait donner lieu à des notes ou à des scores, ce sont les items qu'il conviendrait de pondérer, positivement ou négativement. Dans ce cas, le point délicat concerne la distribution des pondérations sur l'ensemble des items d'un test ou d'un questionnaire, et cette distribution dépend des questions utilisées.

Le problème des notes, qui ne se pose vraiment que dans le cas particulier d'évaluations sommatives, certificatives ou non, ne concerne donc pas directement les questions de la base ; il concerne les épreuves déjà constituées ou envisagées. Des recherches annexes de nature plus docimologique que didactiques pourraient être entreprises pour tenter de trouver des solutions satisfaisantes.

### **Situation d'évaluation**

Dans une évaluation, les questions sont souvent insérées dans des unités plus importantes : épreuves, questionnaires, tests, qui constituent des éléments de la situation d'évaluation.

D'autres éléments interviennent dans la définition des situations d'évaluation : aides et temps mis à la disposition des élèves, proximité des apprentissages, enjeu pour les élèves, nature du curriculum...

## Utilisation de la base et questions de propriété

La base que nous mettons en place ne présentera d'intérêt que si elle rassemble des questions provenant d'évaluations diverses. Pour l'instant les questions présentes dans la base proviennent principalement, par ordre d'importance quantitative, des évaluations de l'APMEP, de celles du SPRESE et de la DEP, et de celles de l'INRP. Quelques questions proviennent de ERSM, de l'IEA, de l'APU (voir ci-dessous).

La base EVAPMIB n'est a priori qu'une base de consultation ; le droit de consulter n'impliquant pas le droit d'utiliser ( nous appelons utilisation des questions leur insertion dans de nouvelles évaluations à quelque niveau que ce soit).

### **Les questions appartiennent aux institutions ou auteurs d'origine.**

Pour une utilisation expérimentale limitée à quelques classes, les questions peuvent être utilisées en précisant l'origine globale EVAPMIB sur les épreuves.

Des négociations sont en cours avec les institutions qui sont à l'origine des questions ; les résultats de ces négociations pourront nous amener à retirer certaines questions de la base si un accord ne peut être trouvé, ou encore à en restreindre l'utilisation.

Nous proposerons qu'en dehors de ce contexte expérimental, limité à quelques chercheurs, l'accès à la base puisse se faire par abonnement. Cela garantira les mises à jours régulières (deux fois par an par exemple) et la diffusion des documents d'accompagnement sur papier.

### **Divers types d'abonnements pourraient coexister :**

**Abonnement professeur** : abonnement individuel permettant l'utilisation de la base au niveau d'une ou plusieurs classes.

**Abonnement établissement et Institutions** : abonnement permettant l'utilisation de la base au niveau de l'établissement concerné. Le coût de l'abonnement dépendrait du type et de la taille de l'établissement.

Des utilisations dépassant le niveau des classes du professeur, des classes de l'établissement ou des actions de formation de l'institution abonnés, seraient soumises à déclaration et des droits supplémentaires seraient perçus.

Une partie du montant des abonnements servirait à assurer le développement permanent et la maintenance de la base, une autre partie serait distribuée en droits d'auteurs proportionnellement à la présence dans la base des institutions concernées.

**L'APMEP** qui, pour l'instant, est propriétaire d'environ 75% des questions, et qui à terme restera propriétaire d'environ la moitié des questions pourrait assurer la gestion financière de la base.

Les lecteurs intéressés peuvent s'adresser à l'IREM de BESANÇON pour obtenir communication d'une **maquette de démonstration** de la base (contenant un petit nombre de questions, et donc pas vraiment opérationnelle).

## Évaluations sources

Sauf précisions, les évaluations ont été faites vers la fin d'année scolaire et portent sur des échantillons représentatifs importants (plusieurs milliers d'élèves).

Toutes les évaluations ci dessous ont produit des rapports et autres documents dont la liste serait trop longue à présenter.

Depuis, plusieurs années, l'IREM de BESANÇON puis l'observatoire EVAPM et la base EVAPMIB entretiennent des rapports suivis avec un certain nombre d'institutions françaises et étrangères qui procèdent à des évaluations à grande échelle. La liste qui suit n'est que provisoire et nous cherchons à élargir ce réseau à la fois dans le temps et dans l'espace géographique.

### **Évaluations françaises (par ordre alphabétique)**

#### **APMEP**

*(Association de Professeurs de Mathématiques de l'Enseignement Public)*

Études EVAPM (Évaluation de Programmes de Mathématiques) menées à partir de 1987 de la Sixième à la Première (en cours).

Ces évaluations se font sur la base du volontariat et portent chaque année sur plus de 100 000 élèves. La représentativité, non évidente a priori, a pu être testée par comparaison avec les évaluations de la DEP. Un biais existe sans doute, mais il est très léger et sans importance pour l'usage prévu de la base.

**EVAPM 6/87 et EVAPM 6/89** : niveau Sixième, 1987 et 1989.

**EVAPM5/88 et EVAPM5/90** : niveau Cinquième, 1988 et 1990.

**EVAPM4/89 et EVAPM4/91** : niveau Quatrième, 1989 et 1991.

**EVAPM3/90 et EVAPM3/92** : niveau Troisième, 1990 et 1992.

**EVAPM2/91** : niveau Seconde, 1991.

**EVAPM1/93** : niveau Première, en cours.

#### **ERSM**

*(Épreuve Standardisée de Raisonnement Mathématique)*

Recherche menée de 1971 à 1974 par J Pelnard-Considerere et J.Levasseur dans le cadre du CNAM (Conservatoire National des arts et Métiers) .

Niveaux CE1 à Quatrième ; échantillons représentatifs (non stratifiés) mais faibles effectifs (200 à 300 par niveaux).

#### **INRP**

Enquête sur l'enseignement des mathématiques à l'école élémentaire.

Fin CE2 - 1976 , Fin CM2 - 1977.

Source importante de questions pour le niveau Sixième.

#### **IREM de BESANÇON**

Il s'agit d'évaluations effectuées dans l'académie de BESANÇON par un groupe de recherche de l'IREM de 1980 à 1983.

Les groupes évalués sont de tailles variables (quelques centaines à quelques milliers) et de représentativité non garantie. En l'absence d'autres sources plus rigoureuses, les questions IREMBES constituent cependant des indicateurs précieux.

**IREMBES 6 ; IREMBES 5 ; IREMBES 4 ; IREMBES 3**

#### **SIGES - SPRESE - DEP**

Il s'agit des appellations successives du service, puis de la Direction, du Ministère de l'Éducation Nationale français, chargé, pour une part, de l'évaluation du système éducatif.

**SIGES 6/80** (évaluation de début d'année)

**SIGES CM2/81**

**SPRESE 5/82 et DEP 5/88** : niveau Cinquième 1982 et 1988

**SPRESE 3/84 et DEP 3/90** : niveau Troisième 1984 et 1990

**SPRESE 2/86** : niveau Seconde 1986

**DEP CE2/89 ; DEP CE2/90 ; DEP CE2/91 ; DEP CE2/92**

**DEP 6/89 ; DEP 6/90 ; DEP 6/91 ; DEP 6/92**

Évaluations organisées à des fins diagnostiques au début de chaque année scolaire, en Sixième et au CE2, depuis septembre 1989.

## Évaluation Nationales d'autres pays

### **Grande - Bretagne (Angleterre, Pays de Galle et Irlande du Nord)**

**APU** (Assessment Program Unit)

Phase 1 : 1978 - 1982 (11ans - 15 ans)

Phase 2 : 1984 - 1988 (11ans - 15 ans)

### **U.S.A.**

**NAEP** (National Assessment of Educational Progress)

1990 - populations grades 4, 8 et 12 (environ CE2, Quatrième, Terminale)

### **SUISSE**

**IRDP** (Institut Romand de Documentation Pédagogique)

Connaissances mathématiques à l'école primaire (en Suisse, l'école primaire se prolonge au niveau de la classe de Sixième française).

Évaluations conduites de 1975 à 1981, de la première à la sixième année de l'enseignement primaire.

## Évaluation internationales

### **Évaluations de l'IEA**

*(International Association for the Évaluation of Educational Achievement)*

**FIMS** : *First International Mathematics Study*

Première enquête internationale sur l'enseignement des mathématiques.

Étude menée au cours des années 1960-64 - (12 pays dont la France).

**SIMS** : *Second International Mathematics Study*

Seconde enquête internationale sur l'enseignement des mathématiques.

Étude menée au cours des années 1980-84. (21 pays dont la France et la Belgique).

Quatre populations : 9 ans, 13 ans, fin d'études secondaires toutes sections, fin d'études secondaires scientifiques.

Rapport français publié par l'INRP.

**TIMSS** : *Third International Mathematics and Science Study*

Troisième enquête internationale sur l'enseignement des mathématiques et des sciences.

Étude en cours (Plus de 50 pays dont la France, la Belgique et la Suisse).

Phase 1 : 1992 - 1995 ; Phase 2 : 1995 - 1999.

Quatre populations : 9 ans, 13 ans, fin d'études secondaires toutes sections, fin d'études secondaires scientifiques.

### **Évaluations de ETS (*Educational Testing Service*)**

**FIAEP** : *First International Assessment of Educational Progress*

Étude menée en 1988 dans six pays (la France ne participant pas). Population : 13 ans.

Questions empruntées à l'évaluation 1986 du NAEP.

**SIAEP** : *Second International Assessment of Educational Progress*

Étude menée en 1991 dans 13 pays dont la France et la SUISSE - 9 ans et 13 ans.

Rapport français publié par la DEP.

# Bibliographie sommaire

## **APMEP (Association des Professeurs de Mathématiques de l'Enseignement Public)**

Brochures EVAPM, contenant les épreuves, les résultats et les analyses des évaluations des programmes de mathématiques menées par l'APMEP (BODIN. A., et All) :

**EVAPM6/87** - Évaluation du programme de Sixième 1987 (Paris 1987)

**EVAPM5/88** - Évaluation du programme de Cinquième 1988 (Paris 1988)

**EVAPM4/89** - Évaluation du programme de Quatrième 1989 (Paris 1989)

**EVAPM3/90** - Évaluation du programme de Troisième 1990 (Paris 1990)

**EVAPM6/89-5/90** - Compléments 1991 des évaluations Sixième et Cinquième

**EVAPM2/91** - Évaluation du programme de Seconde 1991 (Paris 1991)

AUGER. R. ; LAVAULT. D. - Le logiciel MicroCAT : points de vue critique en regard de sa praticabilité pour différentes clientèles

Mesure et Évaluation en Education - Vol 14/3(1991)

BODIN A. (1985) : Problèmes de l'évaluation des savoirs mathématiques

- Petit x n°7/1985 - IREM de Grenoble.

BODIN A.(1992) : Reflexion sur les représentations, les conceptions et les compétences

- Petit x , 30/1992

BODIN A., (1988) : Evaluation in mathematics : the quality of students'knowledge - 6ème congrès international sur l'enseignement des mathématiques (ICME.6). Budapest Août 1988. Traduction française: l'évaluation du savoir mathématique. Bulletin de l'Association des professeurs de mathématiques.(APMEP) N°368/1989 pp 195-219.

BODIN A., (1989) : Rôle de l'évaluateur dans une évaluation à "grande échelle". Actes de la Rencontre 1989 de l'Association pour le développement des méthodologies d'évaluation en éducation, ADMEE-EUROPE Paris Septembre 89 - Publié par l'INRP (L'évaluateur en Révolution - INRP 1990)

BODIN A., (1992) : What does it mean to assess - Conférence ICMI sur le thème " Assessment in mathematics Education and its effects" Colonge (SPAIN) Avril 1991 et ICME 7 - QUEBEC..

CARDINET. J., (1986) : Evaluation scolaire et mesure - De Boeck Université

CHEVALLARD. Y (1985) - Vers une analyse didactique des faits d'évaluation - In L'évaluation : approche descriptive ou prescriptive (actes du colloque ADMEE - bruxelles 1985) - De Boeck Université.

De KETELE J.M.(1980) : Observer pour éduquer - Peter Lang.

DOUADY.R. - Jeux de cadres et dialectique outil-objet- Thèse de doctorat - Université de PARIS VII.

EDUCATIONAL TESTING SERVICE (1989) : A world of differences - An international assesment of mathematics and science

EDUCATIONAL TESTING SERVICE (1991) : Learning Mathematics.

GRAS, R. (1986) : Recherches sur l'apprentissage: Analyse des correspondances et méthodes statistiques apparentées - Cahier du Cirade - Université du Quebec à Montreal.

GRAS.R (1977) : Contributions à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques - Thèse- université de RENNES (1977).

HAMBLETON, R.K. ; SWAMINATHAN, H (1985) : Item Response Theory - Principles and Applications Kluwer. Nijhoff.



**IREM de BESANÇON** - BODIN.A et all.(1981- 1985) : OBJECTIFS et EVALUATION ,  
fascicule 1 : généralités  
fascicule 2 : niveaux 6ème et 5ème.  
fascicule 3 : niveaux 4ème et 3ème.

NATIONAL RESEARCH COUNCIL (U.S.A) (1989) : A Framework for International Comparative Studies in Education.

NUTTALL, D. L. (1987) : The validity of assessments - European Journal of Psychology of Education - 1987 Vol II

ROBIN, D.; BARRIER, E. (1985) : Enquête internationale sur l'enseignement des mathématiques - Le cas Français. - Institut National de Recherches Pédagogiques (Paris).

ROBITAILLE F. , GARDEN A. (1989) : The IEA Study of Mathematics II : Contexts and Outcomes of School Mathematics. Pergamon Press

STAKE, R. (1975) : Evaluating the Arts of Education : A responsive Approach, Colombus, Merrill.

TIMSS (Third International Mathematics and Science Study) (1990) : TIMSS - an introduction.

TOURNEUR Y, ROCHEZ-NIMAL.P (1987).: Macro-évaluation: finalités, objets d'étude et méthodes.- Mesure et évaluation en éducation, Vol 10 - n°3

TOURNEUR.Y, BOUILLON. D., : Les questions en mathématiques : étude critique - Mesure et Evaluation en Education - Vol 12/2/3(1989)

TRAVERS K. J. , WESTBURY I (1989) : The IEA Study of Mathematics I : Analysis of Mathematics Curricula. - Pergamon Press.