

TECHNIQUES SUSCEPTIBLES DE FACILITER LES DIAGNOSTICS INDIVIDUELS ET COLLECTIFS

*Réflexions sur les notions de niveau et d'hétérogénéité
(A propos des évaluations de début d'année organisées par la D.E.P.)*

Antoine BODIN
IREM de Besançon

Journées Nationales des Formateurs académiques en Mathématiques
5 juin 1990 - Cité Universitaire Paris

Cet exposé, préparé pour une rencontre de formateurs en mathématiques, ne concerne pas spécialement les mathématiques. Le même exposé aurait pu être fait, dans le cadre actuel des évaluations de la D.E.P pour des formateurs en français (en modifiant quelques exemples). Plus généralement, il concerne l'évaluation diagnostique, quelle que soit la discipline concernée, lorsque cette évaluation s'effectue dans le cadre de la classe en s'appuyant sur des épreuves passées par tous les élèves et comportant un nombre important de questions.

En préparant cet exposé, j'ai d'abord pensé à l'enseignant, qui, après avoir codé les questionnaires d'évaluation remplis par ses élèves, et après avoir reporté les codes dans les grilles de codage, se retrouvait devant des matrices contenant près de 5000 informations, et cela pour une seule classe. J'ai ensuite pensé aux formateurs, puisqu'il s'agit ici d'une réunion de formateurs. Que pouvons-nous proposer aux enseignants qui puisse les aider à tirer le meilleur parti de ces évaluations ?

Au cours de cet exposé, je m'attacherai à explorer les points suivants :

- Caractère didactique de l'évaluation proposée,
- Pluralité des référents possibles,
- Difficultés liées à l'analyse des résultats d'une classe particulière,
- A propos des notions de niveau et d'hétérogénéité,
- Passage d'une évaluation quantitative à une évaluation qualitative,
- Aspects multidimensionnels de la notion de compétence,
- Vers le diagnostic individuel.

Chacun de ces points mériterait de plus amples développements, et, à mon sens, pourrait faire l'objet de travaux pratiques au cours des stages académiques. Les quatre derniers points sont illustrés ici en utilisant des données issues de l'évaluation 1989¹ (mathématiques classe de sixième). Ils pourront tout aussi bien être illustrés à partir des résultats 1990 du CE₂ ou de sixième, en mathématiques ou en français, et éventuellement en croisant les résultats obtenus dans ces deux disciplines.

¹ Pour partie, données communiquées par Roberte Tomassone. (présentées d'un point de vue global au cours de l'exposé précédent)

1. Caractère didactique de l'évaluation proposée

Les évaluations de début d'année de la DEP n'ont pas pour objectif de permettre le repérage d'insuffisances ou de difficultés générales chez les élèves. Pour cela, les enseignants n'ont nul besoin d'instrumentation particulière. L'objectif poursuivi est plutôt de faciliter un repérage fin et précis des difficultés essentielles, de celles qui constituent des obstacles aux acquisitions ultérieures, de celles qui plaçant d'emblée l'élève dans une situation d'échec risquent de rendre inopérantes les tentatives d'améliorations ultérieures, qu'elles proviennent de l'élève ou de l'enseignant.

Au delà de l'analogie médicale assez simpliste : diagnostic, remédiation... qui a pu faire sourire², nul ne peut nier que tout progrès relatif à l'objectif précisé ci-dessus soit de nature à améliorer la qualité générale de l'enseignement des mathématiques.

Il faut noter que le type d'évaluation proposé conduit à aborder le problème des difficultés scolaires sous un angle résolument didactique. Cet aspect des choses à été parfois contesté, de deux manières plus ou moins contradictoires :

- Pour les uns, les élèves en difficulté seraient le plus souvent des élèves peu motivés, plus ou moins en rupture d'école, et, pour les aider, il vaudrait mieux promouvoir des actions à caractère social ou psycho-social. Les erreurs commises par ces élèves seraient surtout des signes de leur manque d'intérêt; leur étude systématique par l'enseignant ne serait pas de nature à l'aider à prendre des décisions susceptibles d'améliorer les performances de ces élèves.

- Pour les autres, les connaissances concernant le fonctionnement du système didactique seraient encore insuffisantes (*elles le resteront sans doute longtemps!*), pour assurer une identification sans équivoque des difficultés et pour permettre de proposer des situations d'apprentissage efficaces pour tous et dans tous les cas.

Il n'existe certes pas de solution miracle qui permettrait de résoudre inéluctablement toutes les difficultés, mais on voudra bien considérer qu'une attention à la fois rigoureuse et bienveillante portée à l'analyse des démarches erronées des élèves puisse être de nature, tout à la fois, à motiver les moins motivés et à aider à progresser ceux qui ne seraient victimes que de résistances didactiques.

Le but de l'évaluation diagnostique n'est pas de discriminer: on peut toujours mettre en évidence des différences entre les comportements de deux personnes ou entre deux productions, mais cela est sans intérêt dans notre cas. Il ne s'agit pas davantage de juger ou de noter les élèves. Il s'agit simplement (mais est-ce si simple?) d'observer, puis d'analyser leurs résultats et leurs démarches pour leur donner du sens, sens par rapport à un savoir en voie de construction, celui de l'élève, et sens par rapport à un savoir-faire professionnel, celui de l'enseignant.

2. Pluralité des référents possibles

Les instruments et techniques d'observation et d'analyse varient selon la nature des objets qu'il s'agit d'observer. Ainsi, l'astronome n'utilisera pas les mêmes techniques pour étudier une galaxie, un amas d'étoiles ou une étoile particulière. De même, les méthodes d'investigation devront varier selon qu'il s'agira d'étudier

² Antoine Bodin - Quelques remarques à propos de l'opération d'évaluation...
 "Petit X" N°22/1990 - Edité par l'IREM de Grenoble

l'ensemble des élèves d'un niveau scolaire donné, l'ensemble des élèves d'une classe particulière, ou un élève particulier.

L'observation des compétences, à un moment donné, de tous les élèves d'un même niveau scolaire (de façon exhaustive ou par échantillonnage), apporte un ensemble d'informations qu'il est possible d'étudier au **niveau global** (Cf exposé de R. Tomassone). Ces études renseignent autant sur le comportement des questions que sur celui des élèves, elles permettent de repérer des comportements typiques, des comportements rares... Si l'on se limite pas à des calculs de pourcentages, elles permettent d'étudier des *dépendances*, voir des *implications* (statistiques) entre questions. Elles permettent de dégager des tendances, et, dans une certaine mesure, avec un risque d'erreur calculé, de prévoir le comportement dans une situation donnée, d'un élève particulier, connaissant son comportement dans d'autres situations.

L'étude peut aussi être faite au **niveau local**, celui d'une classe (éventuellement d'un groupe de classes ou au niveau d'un établissement...). Il faut alors que l'enseignant se donne des moyens (*ou qu'on lui donne des moyens!*) de lecture et d'interprétation des résultats qui soient de nature à l'aider à prendre des décisions relatives à la gestion de sa classe. Dans ce cas, l'évaluation n'est justifiée que si elle permet à l'enseignant de prendre, sans gaspillage de temps, des décisions dont les effets sur les acquisitions des élèves soient meilleurs que ceux résultant des décisions qu'il prend habituellement en se fiant tantôt à son intuition, tantôt à son savoir-faire.

L'étude peut enfin être faite au **niveau individuel**, celui de l'élève. On se trouve alors au niveau de l'expertise. Il s'agit alors d'atteindre une connaissance aussi complète que possible des modes de fonctionnement, des acquis et des lacunes, d'un élève particulier.

Qu'il s'agisse du niveau local ou du niveau individuel, les observations faites ne peuvent prendre de sens que si elles sont simultanément rapportées à trois **référents** (l'ensemble des observations constitue alors le **référé** et ce référé ne peut prendre du sens que rapporté à un univers de possible ou de souhaitable):

- le référent "**comportement global de la population**"

Dans ce cas on parle habituellement d'évaluation normative, le référent étant en quelque sorte la norme du groupe. Dans la pratique, ce premier référent à tendance à être remplacé par le comportement global d'une ou plusieurs classes particulières, ce qui limite singulièrement l'horizon... d'où l'intérêt d'évaluations portant sur des populations plus importantes.

- le référent "**usages et attentes du système**"

Ce référent, partiellement implicite (et partiellement explicité dans les programmes, manuels, examens ...) conduit, lorsqu'il est explicité et opérationnalisé, à l'évaluation critériée. Ce second référent est bien sûr influencé par le premier, mais il peut exister un décalage important entre les deux.

- le référent "**didactique**"

Ce référent concerne le fonctionnement même du savoir. Ici, les comportements observés prennent du sens non en fonction de ce que d'autres font ou en fonction de ce que d'autres attendent, mais en fonction des problèmes qu'ils permettent ou empêchent de résoudre...

La place manque pour donner des exemples. En général, le référent n'est pas explicite ou reste flou. Des phrases du type "Ils ne savent même pas...", "Ils devraient pourtant savoir...", "Il faudrait que les élèves arrivant en sixième sachent...", particulièrement lorsqu'elles sont énoncées dans le cadre de stages, mériteraient d'être soigneusement analysées et leurs contenus situés par rapport aux référents proposés ci-dessus.

3. Difficultés liées à l'analyse des résultats d'une classe particulière

Les quelque 5000 nombres répartis dans les grilles de codage d'une classe peuvent avoir quelque chose de décourageant pour l'enseignant le mieux disposé.

Évaluation Mathématiques/Sixième										
Résultats d'une classe										
	Som	Calc	Que	Item	Δ0	Δ0	Δ0	Δ0	Δ0	Δ0
TO	ARR	2	8	16	1	1	1	1	1	1
TO	ARR	2	8	17	1	1	1	1	1	1
TO	ARR	2	8	18	1	1	1	1	1	1
TO	ARR	2	9	19	1	1	1	1	1	1
TO	ARR	2	9	20	1	1	1	1	1	1
D	D	2	10	21	1	1	1	1	1	1
D	D	2	10	22	1	1	1	1	1	1
TO	Sous	2	11	23	1	1	1	1	1	1
TO	Sous	2	11	24	1	1	1	1	1	1
TO	Sous	2	11	25	1	1	1	1	1	1
TO	Sous	2	12	26	1	1	1	1	1	1
TO	Sous	2	12	27	1	1	1	1	1	1
D	D	2	13	28	1	1	1	1	1	1
D	D	2	13	29	1	1	1	1	1	1
H1	H1	2	14	30	1	1	1	1	1	1
H1	H1	2	14	31	1	1	1	1	1	1
H1	H1	2	14	32	1	1	1	1	1	1
H2	H2	2	15	33	1	1	1	1	1	1
H2	H2	2	15	34	1	1	1	1	1	1
H2	H2	2	16	35	1	1	1	1	1	1
H2	H2	2	16	36	1	1	1	1	1	1
D	D	2	16e	37	1	1	1	1	1	1
H1	H1	2	17	38	1	1	1	1	1	1
H1	H1	2	17	39	1	1	1	1	1	1
H1	H1	2	17	40	1	1	1	1	1	1
H1	H1	2	18	41	1	1	1	1	1	1
H1	H1	2	18	42	1	1	1	1	1	1
H3	H3	2	19	43	1	1	1	1	1	1
H3	H3	2	19	44	1	1	1	1	1	1

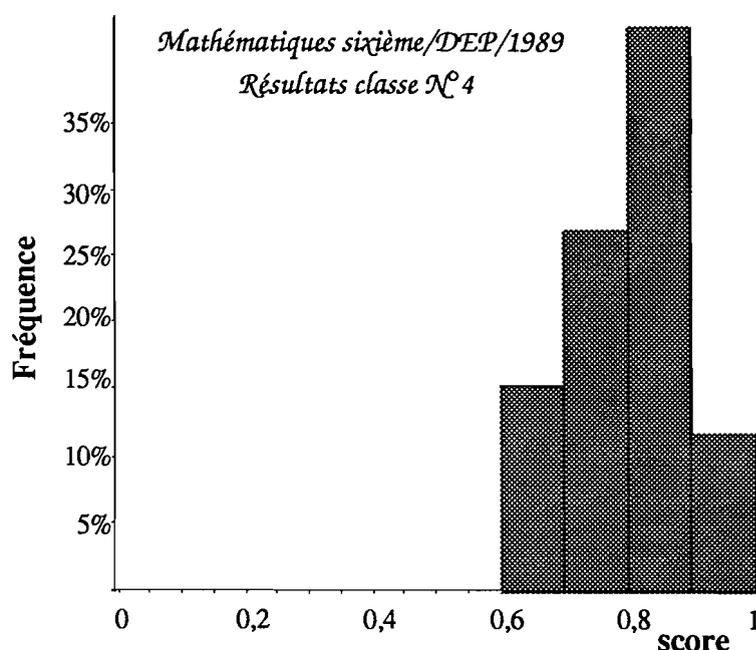
Que peut-il faire de cet amas de chiffres ? S'il s'agit, comme dans ce qui précède, d'en tirer des scores globaux en ne tenant compte que des codes indiquant une réussite, alors, à quoi bon avoir codé différemment des réussites partielles ou divers types d'erreurs ? Mais comment utiliser la richesse de l'information ainsi recueillie sans finir par passer plus de temps à l'analyse des grilles qu'au travail avec les élèves ?

L'utilisation de moyens informatiques semble s'imposer pour organiser, trier, réduire l'information. Il est certain que cette utilisation peut apporter un gain de temps appréciable. Encore faut-il savoir ce que l'on cherche ! Calculer plus rapidement des moyennes, des écarts types, faire des classements d'élèves peut-être utile, mais cela

constitue une sous-utilisation regrettable de l'outil informatique. Pour dire les choses rapidement, il y a même quelque danger à utiliser l'informatique pour ne faire, de façon apparemment plus scientifique, que ce que l'on savait faire sans elle.

Dans ce qui suit, on proposera des pistes qui s'éloignent plus ou moins des pratiques habituelles mais **qui ne requièrent pas formellement l'utilisation de l'informatique**. Une fois les démarches proposées comprises et leur éventuel intérêt reconnu, il sera temps d'avoir recours à l'informatique pour en accélérer le déroulement. Même si l'évaluation dont il est question ici n'a pas pour fonction de discriminer, il est de fait qu'elle discrimine. Que l'on veuille ou non, la première chose qu'elle va révéler est que certains élèves obtiennent globalement de meilleurs résultats que d'autres. Ces différences, qui apparaîtraient quelle que soit la classe, ne sont pas nécessairement stables, c'est à dire qu'elles pourraient apparaître différemment à l'occasion d'évaluations différentes, et ne traduisent pas toujours des différences qualitatives.

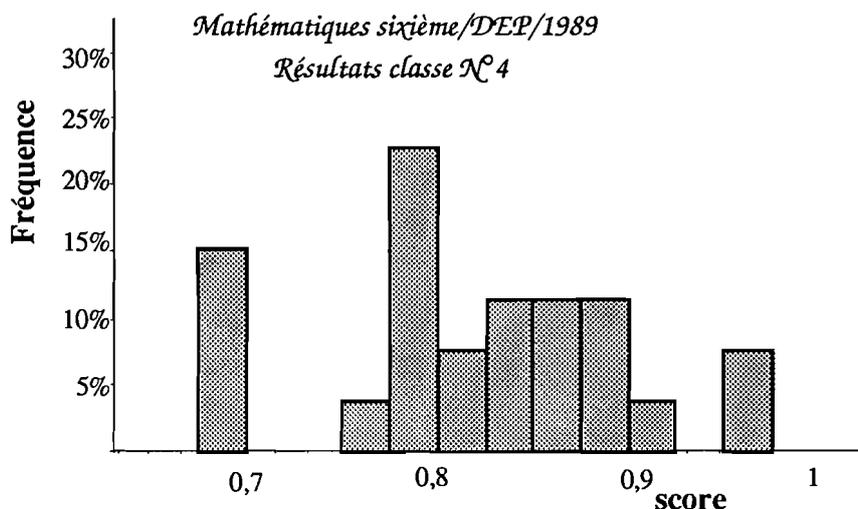
Voici par exemple un graphique représentant les résultats d'une classe particulière (Evaluation DEP 1989 - mathématiques classe de sixième). On a porté en abscisse la proportion de question réussies (scores) et en ordonnée le pourcentage d'élèves de la classe ayant obtenu un tel score. Ainsi, il y a dans cette classe environ 40% des élèves qui ont obtenu un score compris entre 0,8 et 0,9 (80% à 90% de questions réussies). Les élèves correspondants sont au nombre de 11 (sur 26) et le pourcentage n'est indiqué que pour permettre les comparaisons ultérieures.



Que peut penser un professeur devant un tel graphique? Peut-être commencer par se méfier des impressions premières et se poser un certain nombre de questions ! Il ne serait pas surprenant qu'un professeur qui trouverait, pour sa classe, la distribution des scores représentée ci-après soit conduit à envier son collègue titulaire de la classe présentée plus haut. N'a-t-il pas en effet une classe plus homogène et de meilleur niveau moyen ?

En fait il s'agit des **mêmes données** et de **la même classe**. Dans le premier cas, on a simplement utilisé des regroupements d'amplitude plus faible et on a représenté toute l'étendue des scores possibles (0 inclus). On voit ainsi l'intérêt qu'il y

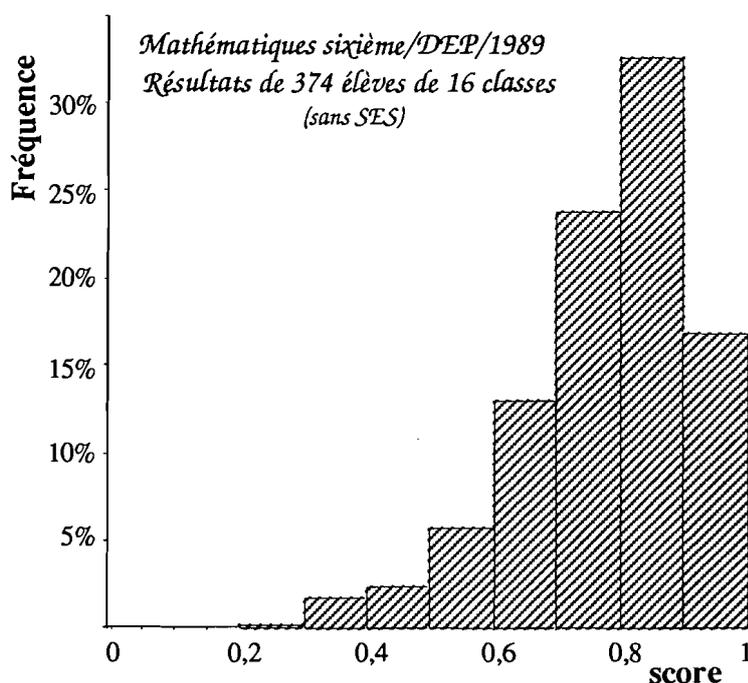
a, lorsque l'on cherche à visualiser les résultats d'une évaluation, à faire figurer sur la représentation les scores minimum et maximum possible.



Cela permet de garder présent à l'esprit que **les élèves faibles savent des chose, parfois même beaucoup de choses**. Qu'ils ne sont faibles que par rapports à d'autres et par rapport aux attentes (légitimes) des enseignants.

L'évaluation fait bien sûr apparaître des différences entre les élèves, mais ces différences sont-elles vraiment importantes ? N'y a-t-il pas parmi les élèves bien placés certains élèves qui commettent des erreurs didactiquement plus importantes que celles commises par certains autres moins bien placés? Les élèves les moins bien placés doivent-ils être considérés comme étant "*en difficulté*" ? L'intérêt de reporter les résultats d'une classe particulière à ceux d'une population plus importante apparaît aussitôt.

Voici un autre graphique qui présente, à la même échelle que le premier, les scores (nombre de questions réussies) à l'évaluation 89 (Mathématiques Sixième) obtenus par les 374 élèves de 16 classes (dont les 23 élèves de la classe précédente).



La comparaison des deux graphiques fait aussitôt apparaître la classe précédente comme une classe ayant un “*bon niveau*” relatif. En fait, elle n’a pas été choisie au hasard, on verra plus loin qu’il s’agit de la classe qui obtient les meilleurs résultats parmi les 17 classes spécialement étudiées pour préparer cet exposé. Cela ne signifie pas pour autant que certains élèves ne soient pas en difficulté sur tel ou tel point particulier, ce que l’analyse qualitative des questionnaires pourra peut-être révéler.

La dissymétrie importante de l’histogramme obtenu indique que, dans l’ensemble de la population, la distribution des scores n’est pas *normale*, en ce sens qu’elle ne suit pas la *loi de Laplace-Gauss*. (ce que le calcul confirme). Contrairement à ce que pensent encore quelques personnes, ce phénomène est à mettre au crédit tout à la fois de la formation antérieure des élèves et de l’évaluation mise en place. La place manque pour justifier ces affirmations. Disons simplement que s’il n’y avait pas formation préalable présentant une certaine homogénéité, ou bien si l’évaluation était faite n’importe comment, on pourrait s’attendre à une distribution *normale* des scores. La réciproque est vraie, au moins pour une évaluation *valide* portant sur un ensemble d’objectifs non indépendants et ayant fait l’objet d’apprentissages spécifiques.

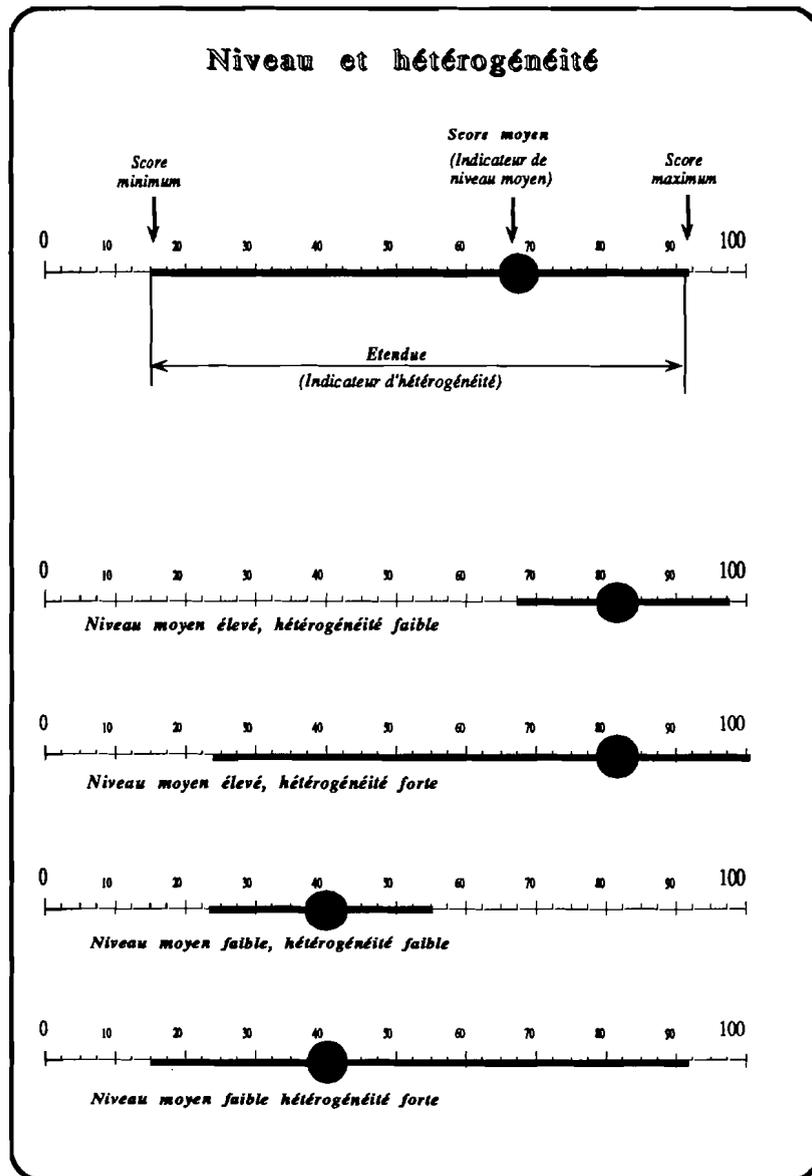
4. A propos des notions de niveau et d’hétérogénéité

On entend souvent les enseignants se plaindre du niveau (faible !) ou de l’hétérogénéité (grande !) de leurs classes. Chacun sait que des classes peuvent avoir des caractéristiques très différentes, mais chacun a tendance à rapporter sa classe à des attentes qui pour être légitimes n’en sont pas moins partiellement arbitraires. On sait en effet que les notions de niveau et d’hétérogénéité, s’ils elles ont un sens, ne peuvent être que des notions relatives.

Pour une classe particulière, on peut, a priori, considérer le score moyen de la classe comme un **indicateur du niveau** de la classe, et l’étendue des scores (écart séparant le meilleur score du score le moins bon) comme un **indicateur d’hétérogénéité** de la classe.

Dans une classe, les élèves ayant les scores extrêmes sont aussi intéressants que les autres, si ce n’est plus. Cela nous conduit à éviter l’utilisation du paramètre classique de mesure de la dispersion: *l’écart type de la série des scores*, pour lui préférer un autre paramètre moins souvent utilisé: *l’étendue de la série des scores*. Le cas particulier, rarement observé, d’un élève à la fois “faible” et très détaché du reste de la classe devrait faire l’objet d’une étude particulière.

La figure ci-après met en évidence les 4 types de classes qu’un enseignant peut s’attendre à rencontrer, qu’elles soient construites de manière délibérée ou qu’elles soient le fruit d’une répartition au hasard. On peut penser que, dans ces divers types de classes, les stratégies d’enseignement les plus favorables aux progrès de l’ensemble des élèves ou aux progrès des élèves les plus en difficulté ne sont pas les mêmes. Par exemple, dans une classe du premier type (niveau élevé, hétérogénéité faible), des travaux en groupes peuvent sans doute être organisés sans précautions particulières en ce qui concerne le mode de groupement des élèves; il n’en serait sans doute pas de même dans une classe du quatrième type (niveau faible, hétérogénéité forte). Il n’est donc pas sans intérêt pour l’enseignant de chercher à rendre plus objectives les impressions qu’il a sur le niveau ou l’hétérogénéité de sa classe.

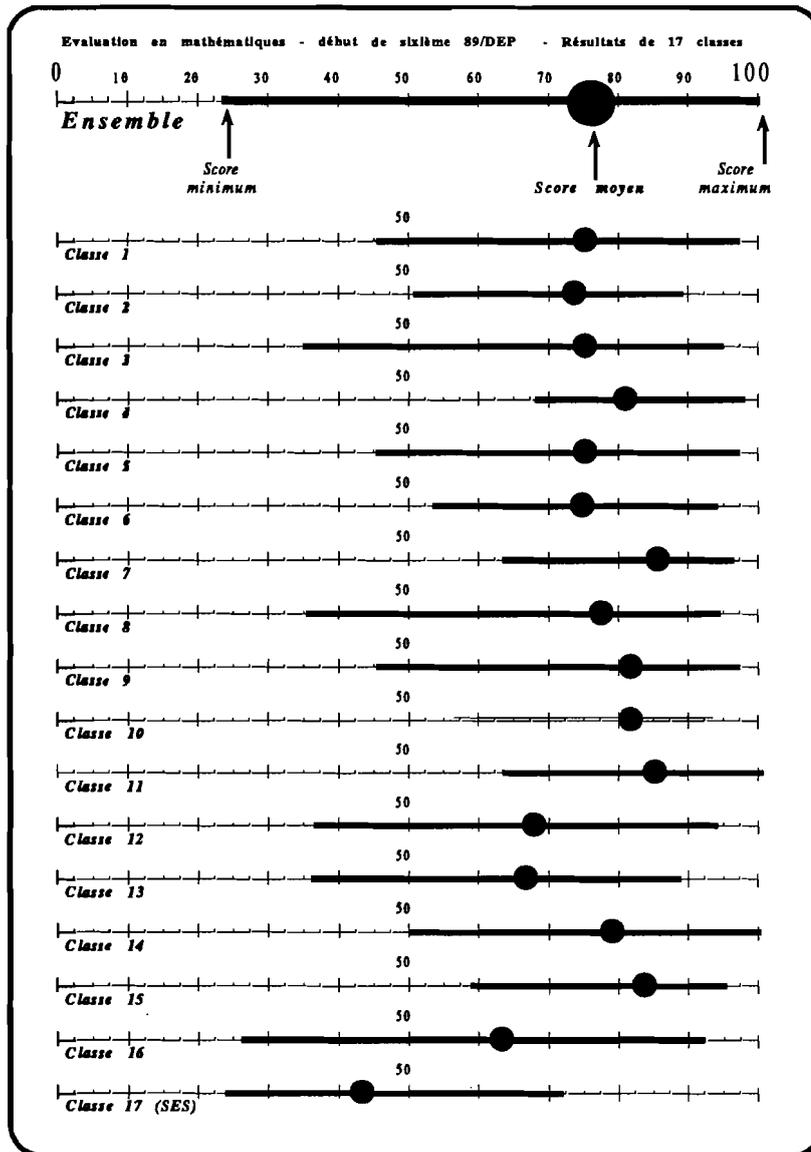


La figure elle, montre comment se situent, en termes de niveau et d'hétérogénéité, les 17 classes que nous avons étudiées. On retrouve, avec tous les intermédiaires, les 4 cas présentés plus haut. Ainsi, les classes n°4 et 9 ont le même score moyen (niveau moyen élevé), mais ont des étendues variant du simple au double. De même, les classe N°16 et 13 ont sensiblement le même score moyen (niveau moyen faible) avec des dispersions assez différentes.

La classe N° 17 est une classe de sixième S.E.S. Dans son ensemble, elle se distingue assez nettement des autres classes. Il convient toutefois de remarquer qu'elle se distingue autant de la classe n°13 (*classe faible et hétérogène*) que cette dernière se distingue de la classe n°7 (*classe forte et homogène*).

Remarquons aussi que certains élèves de la classe de SES ont un score largement supérieur aux scores obtenus par certains élèves des classes de sixième "normale". Mieux, dans la meilleure des 17 classes présentées (classe 4), on trouve au moins un élève qui a moins bien réussi l'ensemble des épreuves que l'un des élèves de SES. On montrera plus loin que, si au lieu de comparer les scores, on compare les réussites dans les différents domaines (techniques opératoires, sens des opérations, etc.), c'est

alors, et de loin, la classe de SES qui doit être considérée comme la plus hétérogène.



Deux élèves peuvent en effet obtenir le même score global à partir de comportement très différents, soit ici, à partir d'ensembles très différents de questions maîtrisées. Il est clair que ce phénomène, impossible en cas de réussite totale (score 100), est de plus en plus possible lorsqu'on descend vers les bas niveaux de réussite.

Le type d'études ci-dessus peut aider l'enseignant à relativiser son jugement quant au caractère éventuellement faible ou hétérogène de sa classe. Le professeur de la classe n°4, par exemple, s'il était resté isolé, aurait pu estimer que sa classe était faible, et, pourquoi pas, hétérogène. Il convient de ne pas oublier que les attentes élevées des enseignants, sont, si elles restent raisonnables, des conditions nécessaires au progrès des élèves. Dans le cas présent, ce professeur sera conscient que sa classe possède des qualités que d'autres n'ont pas, qualités sur lesquelles il pourra s'appuyer, mais aussi qualités qui lui donnent des responsabilités particulières. De son côté, le professeur de la classe n°16 saura que des difficultés l'attendent; il s'attendra à subir des échecs qui

ne seront moins vécus comme étant les siens propres, il aura la fierté de remplir une tâche difficile, sera plus attentif à ses élèves et ressentira comme autant de victoires les obstacles surmontés par ses élèves. (*Il n'est pas dans mon propos d'aborder ici la question des aides à apporter à ce professeur (formation - heures de soutien...), mais il est clair que la question se pose.*)

Il convient aussi de remarquer que si les épreuves d'évaluation mises au point par la D.E.P. permettent de déceler des lacunes particulières et des dysfonctionnements³, elles ne sont pas construites pour mettre en évidence l'ensemble des qualités des élèves. Ainsi, elles ne prétendent rendre compte ni de leurs intérêts ou de leur motivation en général, ni de leur créativité, ni de certaines capacités spécifiques, particulièrement de nature géométriques. Il y a donc lieu d'éviter de tirer de ces évaluations des conclusions définitives que ce soit en ce qui concerne le *niveau* de la classe ou celui d'un élève particulier.

Situer sa classe est ainsi rendu possible d'une façon plus objective qu'habituellement et nous avons vu que cela n'était pas sans intérêt. Toutefois, il est clair que cela donne bien peu d'indications sur ce qu'il conviendrait de faire pour améliorer les compétences de l'ensemble des élèves, et plus particulièrement des élèves reconnus comme étant "*en difficulté*".

5. Passage d'une évaluation quantitative à une évaluation qualitative

Dans tout ce qui précède nous n'avons utilisé que les scores obtenus par les élèves à l'ensemble des épreuves. Pour en arriver là, il nous a fallu laisser de côté une grande partie de l'information recueillie :

- les informations concernant la qualité des réussites, les types d'erreurs et les non-réponses.
- le caractère plus ou moins conjoint ou plus ou moins disjoint, pour un même élève, de certaines types de réponses.

Par exemple, un élève ayant obtenu un score total de 50 réponses correctes (sur 71, soit 77%), peut aussi bien avoir 21 non réponses que 21 erreurs significatives. Ces erreurs peuvent être groupées dans un ou plusieurs domaines particuliers, par exemple *techniques opératoires (TO)* et *connaissance des nombres (N)*. Elles peuvent au contraire être assez uniformément réparties sur l'ensemble des domaines.

Le tableau ci-après donne, pour la classe n°16 (voir plus haut), les scores des élèves de la classe suivant les domaines. On est encore loin de prendre en compte la totalité de l'information, mais nous avons maintenant la possibilité de distinguer des comportements différents et de commencer à penser en termes de *profils*. De nombreuses observations peuvent être faites sur ce tableau qu'il n'était pas possible de faire sur le tableau rassemblant l'ensemble des résultats. Par exemple, les élèves A003 et A17 présentent des profils opposés en ce qui concerne les domaines L et TO.

³ Cf l'étude de F. PLUVINAGE : *Les mathématiques en sixième après l'évaluation nationale*; IREM de Strasbourg

Mathématiques sixième/DEP/1989
Etude d'une classe - Scores par domaines

	A001	A002	A003	A004	A005	A006	A007	A008	A009	A010	A011	A012	A013	A014	A015	A016	A017	A018	A019	A020	A021	A022	A023	MAX
D	14	13	14	13	10	9	12	14	13	10	12	12	12	12	12	13	11	13	13	10	9	13	11	15
L	8	7	10	7	6	4	6	10	10	9	9	5	8	8	10	8	6	7	6	6	9	7	9	10
N	16	19	18	20	19	14	18	19	19	16	20	17	18	18	20	18	14	20	17	12	9	19	20	20
S	6	7	7	6	6	7	8	7	8	7	6	5	8	7	7	7	7	7	6	5	6	7	8	8
TO	18	17	14	17	14	13	16	16	18	12	14	15	16	17	15	17	16	16	18	14	9	17	18	18

Les élèves sont codés A001, A002 etc...

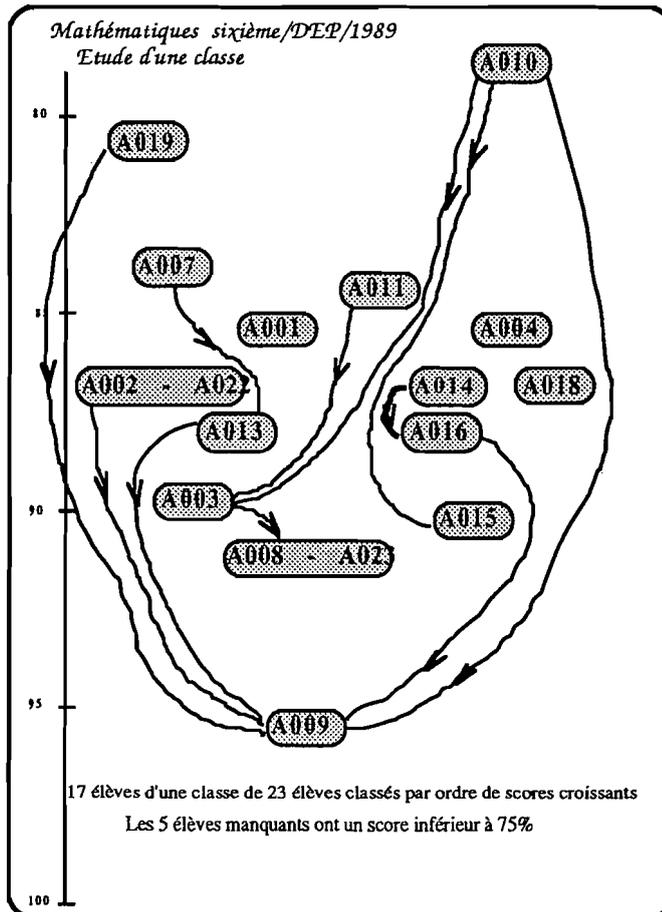
Les domaines sont ceux définis dans les documents de présentation de l'évaluation:

- D : Organisation d'une démarche
- L : Lecture et exploitation des données
- N : Nombres (connaissance des nombres)
- S : Sens des opérations
- TO : Techniques opératoires

Ce tableau montre en particulier l'inconvénient qu'il y a à parler de *niveau*, surtout lorsque l'objectif de l'évaluation n'est pas de *classer* les élèves.

Un niveau absolu en mathématiques, au moins pour une série d'épreuves données, supposerait que pour deux élèves quelconques, on soit assuré que l'un des deux (élève A) ait en particulier réussi toutes les questions réussies par l'autre (élève B). On pourrait alors dire que, par rapport à la situation d'évaluation mise en place, et dont on pourrait encore mettre en doute la validité, le niveau de A serait supérieur au niveau de B. En d'autres termes, nous aurions une *relation d'ordre total* sur l'ensemble des élèves.

Si l'on applique ce raisonnement à la classe N°16, on est amené à constater que le graphe de la relation "avoir un meilleur niveau que" est vide dans cette classe. Même l'élève le moins bien classé pour le score total (A021) a réussi des questions auxquelles l'élève le mieux placé (A009) a échoué.

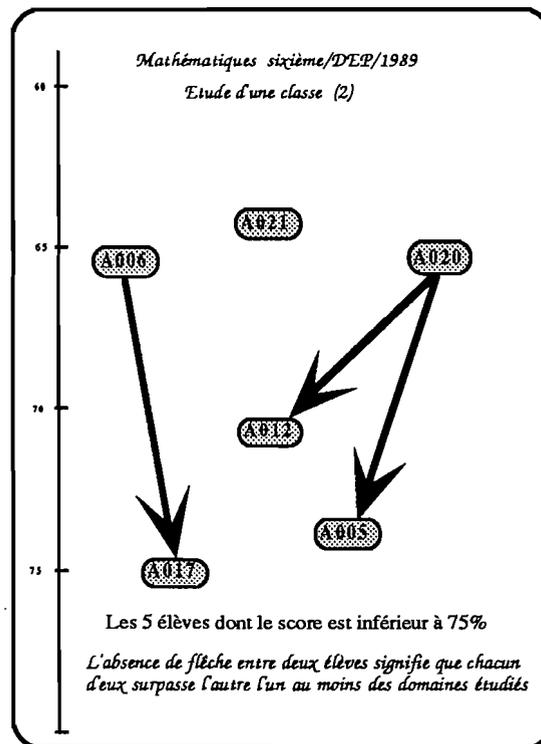


Dans l'évaluation étudiée, les domaines tels qu'ils sont définis dans le document de présentation présentent simultanément une bonne cohérence interne⁴ et une différenciation importante. Il est donc possible de les utiliser pour définir et comparer des profils.

Dans un premier temps, nous avons simplement continué à essayer de donner un sens à la notion de *niveau*. Un élève A est considéré comme ayant un meilleur niveau que l'élève B s'il a obtenu un meilleur score que B dans chacun des cinq domaines considérés. Dans ce cas, dans le graphique ci-contre, on aura lié B à A par une flèche (avec transitivité). Ainsi, dans le sens ainsi défini, l'élève A003 a un meilleur niveau que l'élève A011 (de même que A008 et A023), mais il n'a pas un meilleur niveau que l'élève A019 qui obtient pourtant un score total très inférieur.

Sur le graphique, l'axe vertical repère les scores totaux des élèves. Ainsi, A010 a obtenu un score de 77% tandis que A009 a obtenu un score de 96%. Les élèves A002, A022, A014 et A018 ont obtenu le même score (88%) mais seuls A002 et A022 présentent des profils identiques.

Pour limiter l'enchevêtrement des flèches, le graphique ci-dessus ne prend en compte que les 17 élèves les mieux placés. Les cinq élèves restants sont représentés dans le graphique suivant.



On aurait pu limiter ce diagramme aux trois élèves ayant obtenu les scores les plus faibles (A021, A006 et A020). Dans ce cas il n'y aurait aucune flèche.

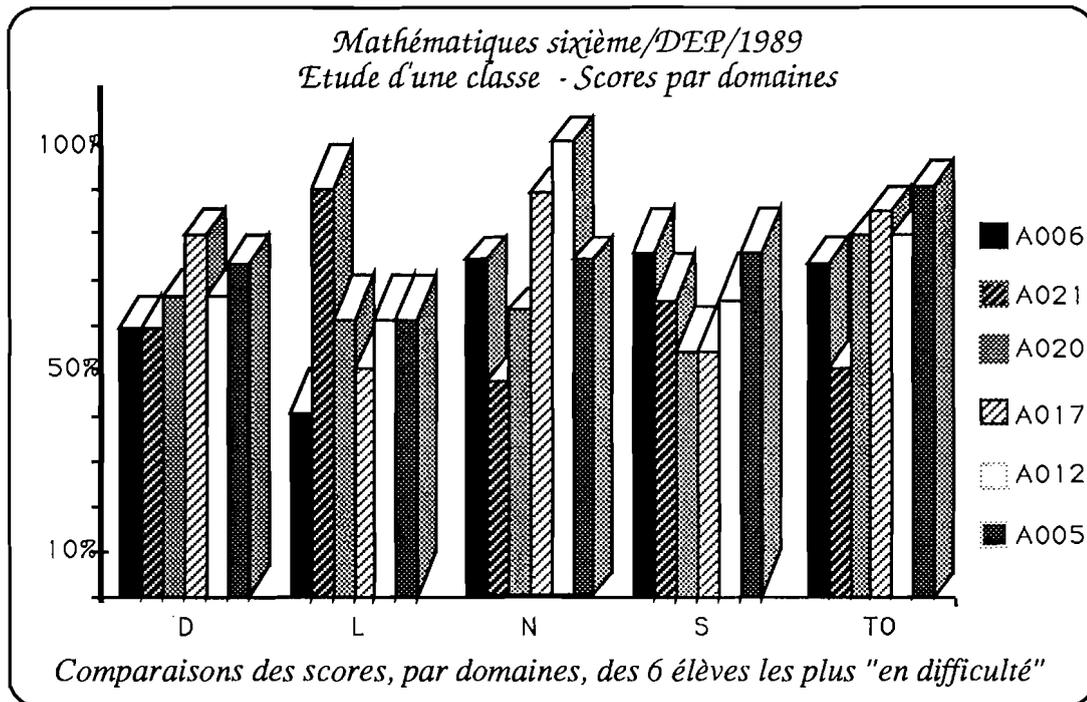
Ce dernier point, déjà signalé dans le cas de la sixième de S.E.S est général : **l'hétérogénéité maximum est obtenue pour des élèves globalement faibles.** Ces élèves sont en difficulté pour des raisons différentes, et un même score global faible peut masquer des différences très importantes.

⁴ Voir l'étude de R. Tomassone : Evaluation sixième - Analyse multidimensionnelle des résultats.

L'évaluation n'aurait-elle conduit qu'à une meilleure analyse et prise en compte des différences existant entre les compétences des trois élèves signalés ci-dessus qu'elle serait déjà justifiée.

Le graphique ci-après permet de visualiser les scores, dans les domaines étudiés, des six élèves qui font l'objet du diagramme de gauche.

On voit sans peine que chacun de ces six élèves manifeste, dans l'un au moins des cinq domaines, une maîtrise, que l'on pourrait qualifier de correcte. De plus, cinq d'entre eux sont supérieurs à chacun des cinq autres dans au moins un domaine.



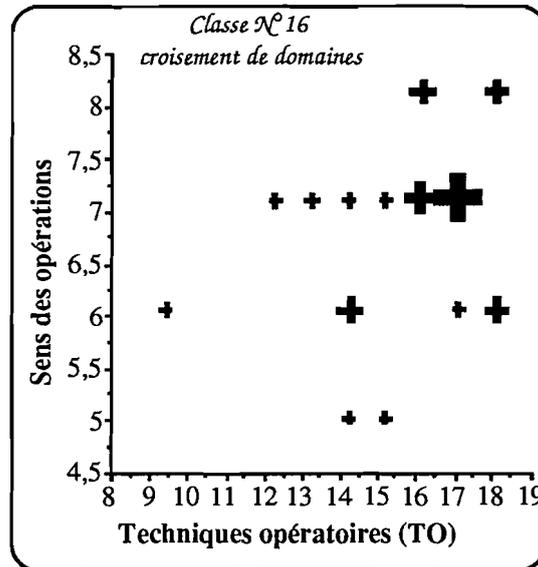
La plupart des propos qui précèdent étaient essentiellement destinés à mettre en cause les notions de niveau et d'hétérogénéité, qui, bien que largement dénuées de consistance, continuent à servir de référence à beaucoup d'enseignants, les empêchant de s'investir résolument dans ce qui pourrait être une véritable *évaluation diagnostique et formative*, au service des apprentissages.

Il est temps maintenant de voir comment compléter l'analyse pour exploiter au mieux les informations recueillies.

6. Aspects multidimensionnels de la notion de compétence

Ce qui précède montre bien qu'il est difficile, sinon impossible, de définir la notion de niveau de façon absolue. De plus, cela serait sans aucun intérêt pour aider les élèves à progresser. Il convient bien au contraire de reconnaître une fois pour toutes que les capacités sont multiformes, que tel élève qui n'a pas encore réussi à *fixer* les tables de multiplication ou qui ne maîtrise pas l'algorithme de la soustraction est capable de venir à bout d'une situation complexe mettant éventuellement en jeu des nombres, en mettant lui même au point des démarches efficaces.

Les modes de pensée arithmétique, algébrique, géométrique, analytique, ...ne sont pas semblables et plusieurs études montrent que les compétences entre ces divers domaines sont faiblement corrélées. Les cinq domaines définis pour l'évaluation en mathématiques sont suffisamment homogènes pour qu'il soit intéressant de croiser les scores obtenus dans ces divers domaines. Par exemple, le graphique suivant croise les scores obtenus par les élèves de la classe n°16 dans les deux domaines **Techniques opératoires** et **Sens des opérations**. Chaque petite croix représente un élève pour lequel on peut lire son score **TO** en abscisse et son score **S** en ordonnée. Les croix plus grosse signalent, selon leurs tailles, la superposition de 2, 3, 4..élèves ayant obtenus les mêmes scores dans les deux domaines.



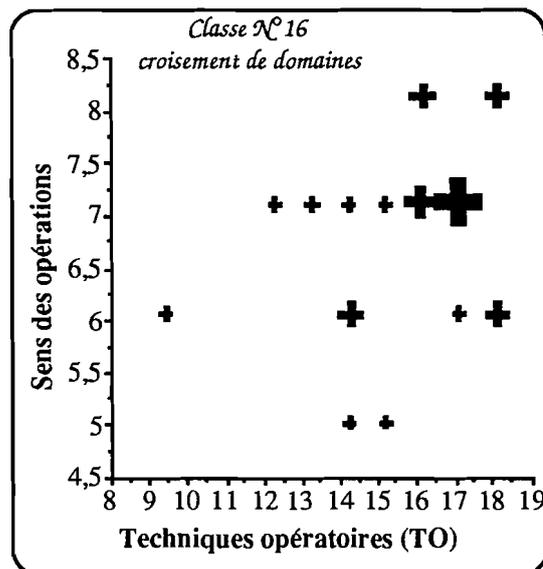
Considérons par exemple les six élèves qui ont obtenu un score de 6 (sur 8) dans le domaine S (ils sont 6). Leurs scores dans le domaine **TO** s'étale de 9 à 18 (sur 18), ce qui correspond sans aucun doute à des niveaux de maîtrise très différents en ce qui concerne les techniques opératoires.

Il est certain que des observations de ce genre peuvent être précieuses pour organiser le groupement des élèves ou choisir le type d'exercice à leur proposer, de façon différenciée ou éventuellement individualisée. Cet exposé se borne cependant à proposer des méthodes d'analyses facilitant le diagnostic, il n'est pas dans mon propos d'aborder la question des décisions pédagogiques qui devront suivre. Par exemple, convient-il de s'appuyer sur les points forts de l'élève pour le faire avancer, de façon détournée, dans les domaines où il est moins à l'aise ? Est-il préférable de l'amener à concentrer ses efforts, directement, sur ses points faibles ? La réponse dépend sans doute en partie de l'élève, mais on le voit bien, le diagnostic n'est qu'une étape.

A partir du tableau du § 5, il est très facile de faire des représentations graphiques du type précédent. Cela peut se faire simplement avec du papier quadrillé et un crayon. Certains enseignants suggèrent de faire réaliser de telles représentations par les élèves eux-même. Outre le fait que ce type de travail fait partie des activités envisageables dans le cadre du programme (gestion de données), il pourrait permettre aux élèves de prendre davantage conscience de l'intérêt de l'évaluation et du type d'exploitation qu'en fait le professeur. De plus, il est remarquable que, répété sur les différents couples de domaines, ces représentations apportent des informations aussi bien sur les élèves *faibles* que sur les *meilleurs* élèves (à condition bien sûr qu'ils n'aient pas obtenu un score total de 100%).

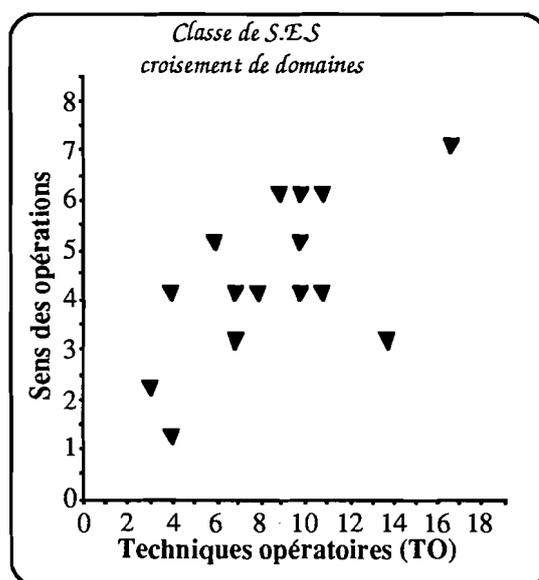
Ici encore, il convient d'être prudents dans l'interprétation des graphiques et, comme dans le cas des histogrammes du § 3, d'utiliser toute l'échelle des scores. La représentation suivante est identique à la précédente, à ceci près que l'on a placé le point (O ; O).

Ce second graphique relativise fortement l'impression d'hétérogénéité, voire de niveau faible que l'on pouvait avoir en regardant le premier.



Le graphique ci-dessous représente le même croisement de domaines effectué pour une classe de S.E.S. Il s'agit d'une classe prise au hasard, donc non représentative, mais ne présentant pas de caractères trop particuliers. Cette fois, et surtout par rapport à la classe précédente, l'hétérogénéité apparaît de façon frappante.

D'une certaine façon, on peut même dire, qu'ici, l'hétérogénéité est maximale: en effet, toute l'échelle des scores est pratiquement couverte, dans chacun des deux domaines. Dans la classe de SES il y a une dépendance assez forte entre les scores obtenus dans les deux domaines. Il y a toutefois un élève qui réussit assez bien en TO (14 sur 18) et qui semble avoir beaucoup de difficultés dans le domaine S.



Il y a 5 domaines, soit, a priori 20 croisements possibles. En fait, cela se réduit à 10 si l'on observe que les domaines ne sont pas ordonnés. Le lecteur trouvera plus

loin, dans le cas de la classe n°16, les 10 graphiques correspondants avec le repérage de quelques élèves seulement pour éviter la surcharge du document.

Tant qu'il s'agit de gestion de l'ensemble de la classe, l'observation de ces graphiques peut suffire pour créer selon les décisions pédagogiques prises, des groupes plutôt homogènes, des groupes hétérogènes, des groupes équilibrés (aide, besoin), bref, des groupes dont la composition, pensée en fonction du projet pédagogique, utilisera pleinement les connaissances acquises sur la variété des compétences des élèves.

Arrivé à ce niveau de l'analyse, **le professeur de la classe aura déjà, et depuis longtemps, une bonne connaissance du contenu de l'évaluation et des types d'erreurs et de réussites des élèves de sa classe.** Pour affiner le diagnostic collectif, il lui faudra sans doute revenir, pour tel ou tel élève, à l'examen des cahiers de tests, mais dans l'ensemble, il lui suffira de revenir à la feuille de codage par domaines. L'utilisation de "surligneurs" permettra alors de mettre en évidence, facilement et rapidement, les questions les moins bien réussies et le type d'erreurs commises.

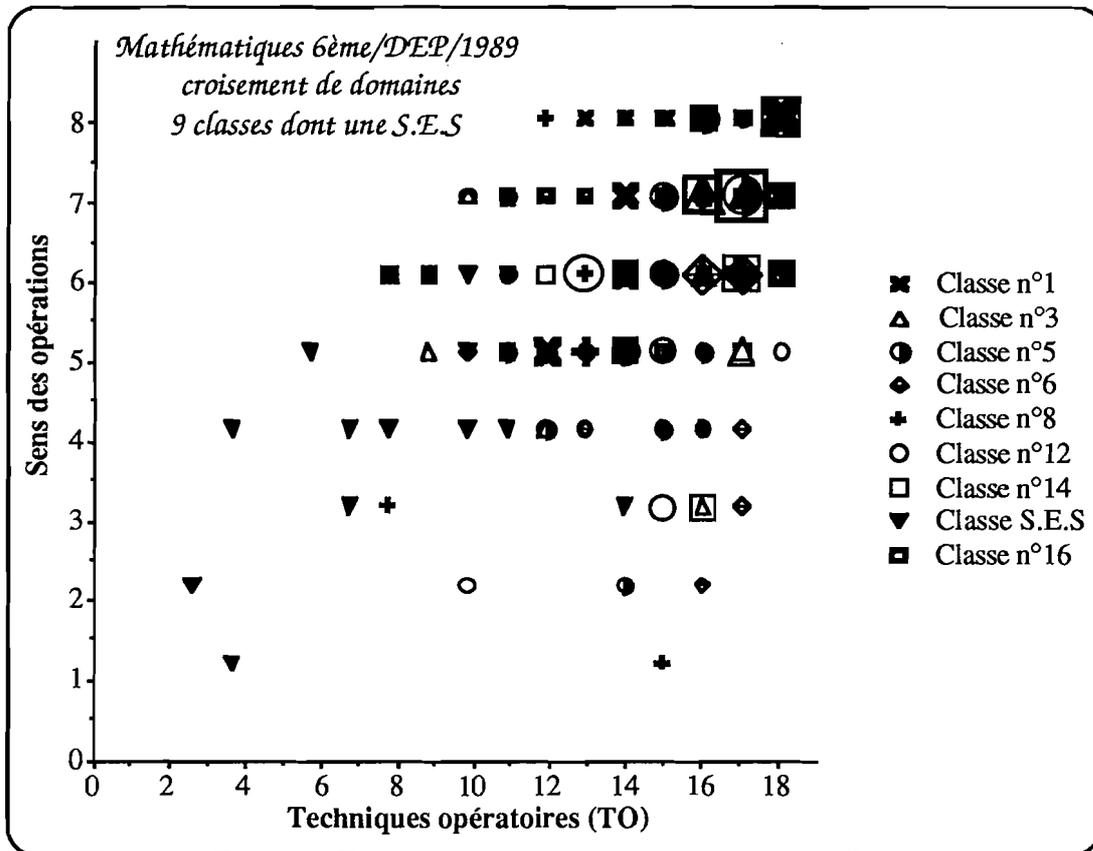
C'est à ce moment que l'informatique devient très utile (mais non indispensable). En effet, après que la saisie des résultats ait été faite une première fois, en suivant l'ordre des questions et des cahiers de tests, avec les codes de réussite, d'erreur, de non-réponse,... On se retrouve avec une grille difficile à utiliser. Il est important d'avoir aussi une grille des résultats par domaines. Il serait aussi utile d'avoir une grille ne comportant que les codes d'erreur...

Grâce à l'utilisation d'un *tableur* et d'un savoir-faire tout à fait élémentaire, il est possible de passer d'une présentation des résultats à une autre en quelques secondes et d'obtenir ainsi toutes les présentations souhaitées. A titre indicatif, on peut estimer que, pour l'opération 6ème/1989, le temps passé au codage aurait été réduit de moitié par une telle utilisation de l'informatique. *Compte tenu de l'importance de l'opération, le gain de temps réalisé équivaldrait au travail annuel de 80 professeurs de mathématiques travaillant 40 heures par semaine!*

On trouvera à la fin de cet article quelques indications sur le matériel et les logiciels utilisés.

Avant de poursuivre vers le diagnostic individuel, et pour illustrer tout à la fois la variété des comportements des élèves et la puissance des moyens informatiques, on trouvera ci dessous un graphique représentant les élèves de 9 classes pour lesquels on a, comme précédemment, croisé les scores dans les domaines TO et S.

Les élèves des différentes classes sont repérés par des symboles différents. On voit en particulier qu'il n'y a pas solution de continuité entre la classe de SES et les autres classes.

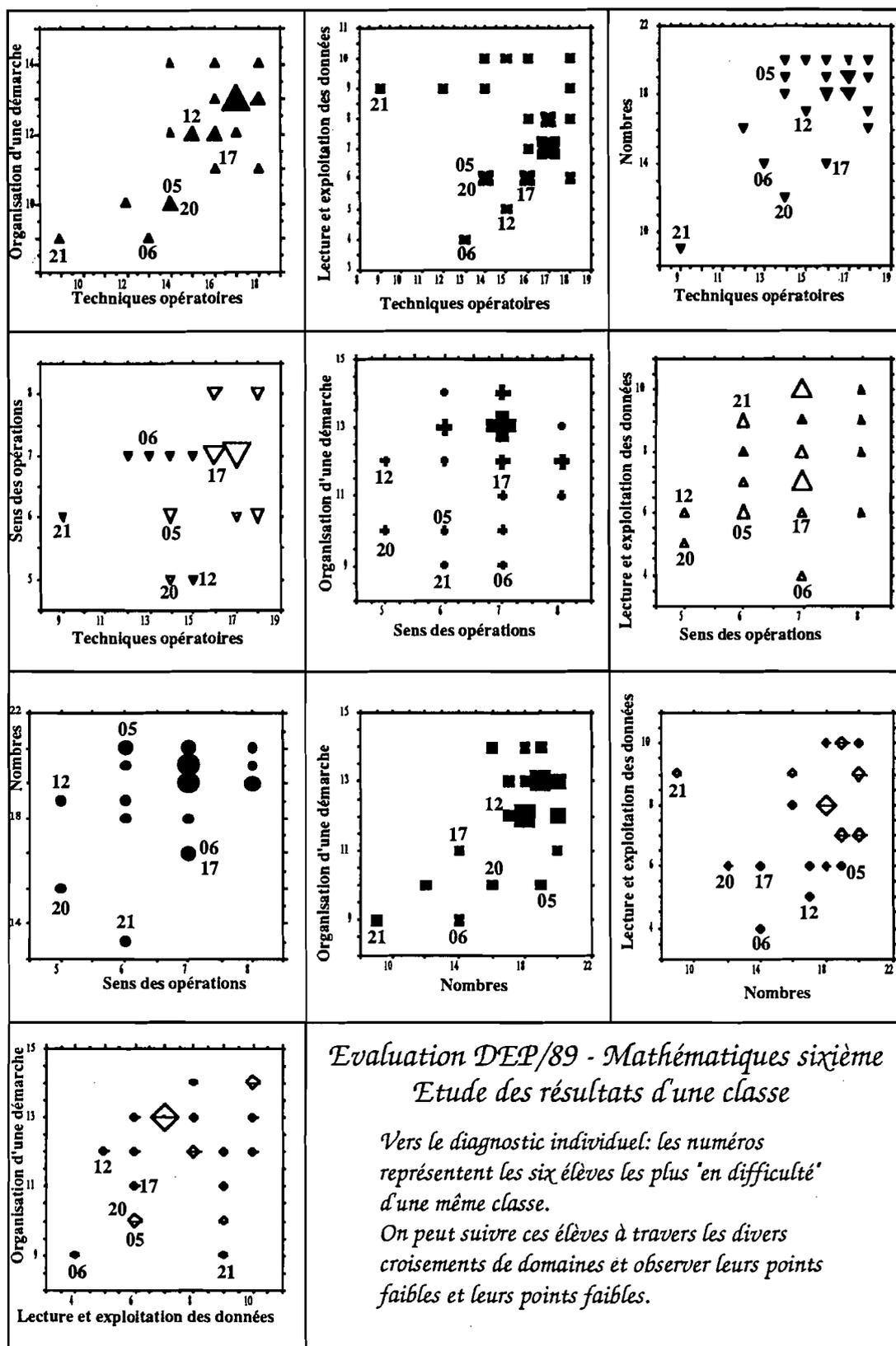


7. Vers le diagnostic individuel

La page suivante présente les croisements de domaines pour tous les cas possibles et pour tous les élèves de la classe n°16. Pour ne pas trop alourdir les représentations, nous avons choisi de ne faire figurer sur les graphiques que les 6 élèves ayant obtenu les scores les plus faibles à l'ensemble de l'évaluation et dont nous avons vu qu'ils se séparaient assez nettement du restant de la classe. Les codes d'identification des élèves (A001, ..., A023) ont été simplifiés et deviennent 01, ..., 23.

Pour faciliter la lecture, on a omis de représenter le point (O;O) sur ces graphiques. Il convient de garder à l'esprit les remarques faites plus haut sur le caractère relativement artificiel de la dispersion mise en évidence.

Le diagnostic individuel suppose le repérage préalable des élèves qui ont besoin de ce diagnostic. Il est raisonnable de commencer par les élèves qui paraissent le plus "en difficulté". Toutefois, les graphiques font apparaître des élèves dont le score moyen est tout à fait honorable mais qui présentent des scores anormalement bas dans un domaine particulier. L'action à mener avec ces élèves sera peut-être plus facile à décider que pour les 6 élèves retenus ici, mais dans tous les cas, l'analyse des croisements aura donné des indications utiles.



L'analyse des graphiques, pour ces six élèves fait apparaître des profils très différents:

L'élève **21**, qui a le score total le plus bas de la classe, obtient de bons résultats dans le domaine **L** (lecture et exploitations de données) et des résultats faibles partout ailleurs (par rapport aux autres élèves de sa classe). Il serait intéressant de comparer avec les résultats obtenus en français par cet élèves et peut être trouver par là un biais pour le réconcilier avec les nombres.

L'élève **05** présente un profil quasiment inversé: il obtient des résultats bons ou corrects dans les domaines **TO** (techniques opératoires) et **N** (connaissance des nombres) et faibles dans les autres domaines.

L'élève **20** est le seul qui ne présente pas de point fort (pour cette évaluation) (voir aussi page 11). Il peut y avoir là une trace d'un manque d'intérêt général pour les mathématiques, peut-être même pour l'école. Il peut aussi s'agir d'une difficulté passagère d'adaptation à un nouveau style d'enseignement.

L'élève **06** a un seul point fort: le domaine **S**. Peut-être que l'utilisation de la calculatrice lui permettra de résoudre des problèmes non triviaux qui lui donneront du goût pour le reste?

L'élève **12** obtient des résultats corrects dans trois domaines: **N**, **TO** et **D**, des résultats plus faibles ailleurs.

L'élève **17** a deux points forts: **TO** et **S**.

Dans tous les cas ci-dessus, il semble que le diagnostic puisse être mis en route dans de bonnes conditions. Pour affiner ce diagnostic, il sera en particulier nécessaire de revenir directement à l'analyse directe de telle ou telle partie des cahiers de tests remplis par ces élèves. Il faudra parfois compléter l'investigation par des questions supplémentaires spécialement choisies pour améliorer notre connaissances sur les compétences et le mode de fonctionnement de tel ou tel élève. Pour les élèves "en difficulté" le diagnostic ne peut être qu'un processus permanent.

Bien sûr, il faudra faire des hypothèses, comme celle des qualités "littéraires" de l'élève **21**, et il faudra éprouver ces hypothèses. D'une façon générale, les conclusions du type de celles qui précèdent demandent à être mise à l'épreuve. Les observations ultérieures peuvent venir infirmer ou confirmer certaines des ces hypothèses ou conclusions. On n'oubliera pas que, de toutes façons, l'objet même de la formation dispensée par l'enseignant est de détruire les indices (négatifs) obtenus lors de l'évaluation.

L'objet de cet article était de mener une réflexion sur le type d'exploitation que l'on pouvait faire, **dans le cadre de la classe**, de résultats obtenus lors d'une évaluation de masse (type "*survey*"), en dépassant les impressions premières que l'on pouvait avoir en ce qui concerne les notions de *niveau* et d'*hétérogénéité*. Il semble que, au cours de l'exposé, nous ayons avancé dans la voie de ce que pourrait être un diagnostic utile, mais il est clair que l'essentiel reste à faire, et **l'essentiel est de l'ordre de la didactique** proprement dite: donner du sens aux erreurs et mettre au point des situations d'apprentissages qui ne nient pas les savoir préalables de l'élève, qui ne renforcent pas ses erreurs et qui lui permettent de progresser de façon significative.

A propos de l'informatique

La saisie des résultats peut être effectuée sur un *tableur* (feuille de calcul). Cette saisie ne demande pas plus de temps que la transcription manuelle des codes sur une grille de codage. Ensuite, il est facile de passer directement d'une présentation des données à une autre: regroupement par domaine, mise en évidence des seuls codes 1, mise en évidence des codes d'erreur, scores par domaines etc...

Nous avons déjà écrit plus haut que l'on pouvait estimer que, pour l'opération 6ème/1989, le temps passé au codage aurait été réduit de moitié par une telle utilisation de l'informatique. *Compte tenu de l'importance de l'opération, le gain de temps réalisé équivaldrait au travail annuel de 80 professeurs de mathématiques travaillant 40 heures par semaine!*

Les traitements statistiques et les représentations graphiques présentées dans ce document ont été faites sur matériel Macintosh en utilisant les logiciels suivants :

EXEL (tableur utilisé pour les calculs)

STATVIEW II (logiciel de traitements statistiques utilisé pour les croisements)

GRAPH (qui permet de réaliser des représentations *en relief*)

On trouve les mêmes possibilités sur matériel PC. Par exemple EXEL est disponible dans beaucoup de collèges et plusieurs logiciels courants permettent de faire des histogrammes et autres représentations graphiques.

Il a été proposé, au cours de l'exposé, de mettre à la disposition des formateurs académiques certains éléments susceptibles de faciliter une introduction prudente de l'informatique dans les stages. Nous pensions en effet qu'au cours de certains stages, il devrait être possible de traiter les données des stagiaires et donc de rendre l'ensemble de la formation plus concrète et plus proche des préoccupations des stagiaires. Bien sûr, cela suppose que les formateurs soient un peu familiarisés avec les techniques et le matériel et qu'ils aient accès au matériel nécessaire. Cela est généralement le cas dans les IREM, c'est aussi le cas dans un certain nombre d'établissements. Dans un premier temps, nous pensions surtout communiquer des feuilles de calcul contenant les résultats des 17 classes étudiées pour préparer cet exposé. En effet, il n'est pas toujours facile de trouver des données prêtes à être exploitées. Ensuite, il nous semblait que les formateurs pouvaient utiliser ces données pour simuler les divers modes de calculs et de présentation de résultats proposés dans ce texte.

Comme annoncé, les données utilisées ici sont communicables, ainsi que moyens de passer d'une présentation à une autre, mais leur intérêt est assez limité du fait qu'il s'agit de données 89.

Nous avons aussi envisagé de proposer un logiciel simple permettant de faire de l'*analyse factorielle des correspondances*, ce qui, dans le cas d'une classe, reviendrait en quelque sorte à faire en une seule fois les croisements que nous avons proposé en dix tableaux. A la réflexion, le logiciel doit encore être amélioré pour être vraiment utile, mais surtout une formation préalable sera nécessaire à son utilisation, en effet, les risques d'erreurs dans l'interprétation des résultats de ce type d'analyse sont importantes. Cela pourrait faire l'objet d'une présentation ultérieures.

La réflexion se poursuit dans ce domaine, et, à terme, il devrait être possible de fournir aux formateurs et aux enseignants des instruments et des techniques de nature à permettre un gain de temps appréciable et une efficacité accrue.