ULB UNIVERSITÉ LIBRE DE BRUXELLES

DI·fusion

Dépôt Institutionnel de l'Université libre de Bruxelles /
Université libre de Bruxelles Institutional Repository
Thèse de doctorat/ PhD Thesis

Citation APA:

Lebrun, B. (1977). *Une évaluation informative* (Unpublished doctoral dissertation). Université libre de Bruxelles, Faculté des Sciences psychologiques et de l'éducation, Bruxelles.

Disponible à / Available at permalink : https://dipot.ulb.ac.be/dspace/bitstream/2013/214257/1/95b7a56f-f8c4-42c9-924e-9aee838b4a26.txt

(English version below)

Cette thèse de doctorat a été numérisée par l'Université libre de Bruxelles. L'auteur qui s'opposerait à sa mise en ligne dans DI-fusion est invité à prendre contact avec l'Université (di-fusion@ulb.be).

Dans le cas où une version électronique native de la thèse existe, l'Université ne peut garantir que la présente version numérisée soit identique à la version électronique native, ni qu'elle soit la version officielle définitive de la thèse.

DI-fusion, le Dépôt Institutionnel de l'Université libre de Bruxelles, recueille la production scientifique de l'Université, mise à disposition en libre accès autant que possible. Les œuvres accessibles dans DI-fusion sont protégées par la législation belge relative aux droits d'auteur et aux droits voisins. Toute personne peut, sans avoir à demander l'autorisation de l'auteur ou de l'ayant-droit, à des fins d'usage privé ou à des fins d'illustration de l'enseignement ou de recherche scientifique, dans la mesure justifiée par le but non lucratif poursuivi, lire, télécharger ou reproduire sur papier ou sur tout autre support, les articles ou des fragments d'autres œuvres, disponibles dans DI-fusion, pour autant que :

- Le nom des auteurs, le titre et la référence bibliographique complète soient cités;
- L'identifiant unique attribué aux métadonnées dans DI-fusion (permalink) soit indiqué;
- Le contenu ne soit pas modifié.

L'œuvre ne peut être stockée dans une autre base de données dans le but d'y donner accès ; l'identifiant unique (permalink) indiqué ci-dessus doit toujours être utilisé pour donner accès à l'œuvre. Toute autre utilisation non mentionnée ci-dessus nécessite l'autorisation de l'auteur de l'œuvre ou de l'ayant droit.

---------- **English Version** ----------

This Ph.D. thesis has been digitized by Université libre de Bruxelles. The author who would disagree on its online availability in DI-fusion is invited to contact the University (di-fusion@ulb.be).

If a native electronic version of the thesis exists, the University can guarantee neither that the present digitized version is identical to the native electronic version, nor that it is the definitive official version of the thesis.

DI-fusion is the Institutional Repository of Université libre de Bruxelles; it collects the research output of the University, available on open access as much as possible. The works included in DI-fusion are protected by the Belgian legislation relating to authors' rights and neighbouring rights. Any user may, without prior permission from the authors or copyright owners, for private usage or for educational or scientific research purposes, to the extent justified by the non-profit activity, read, download or reproduce on paper or on any other media, the articles or fragments of other works, available in DI-fusion, provided:

- The authors, title and full bibliographic details are credited in any copy;
- The unique identifier (permalink) for the original metadata page in DI-fusion is indicated;
- The content is not changed in any way.

It is not permitted to store the work in another database in order to provide access to it; the unique identifier (permalink) indicated above must always be used to provide access to the work. Any other use not mentioned above requires the authors' or copyright owners' permission.

UNIVERSITE LIBRE DE BRUXELLES

Faculté des Sciences Psychologiques et Pédagogiques

UNE EVALUATION INFORMATIVE

Essai de mise au point de méthodes susceptibles de rendre compte d'effets pédagogiques collectifs.

Dissertation présentée pour
l'obtention du titre de
Docteur en Sciences Pédagogiques

par

Bernard LEBRUN

Directeur : Professeur L. VANDEVELDE

OCTOBRE 1977

14514

UNIVERSITE LIBRE DE BRUXELLES

Faculté des Sciences Psychologiques et Pédagogiques

UNE EVALUATION INFORMATIVE

****

Essai de mise au point de méthodes susceptibles de rendre compte d'effets pédagogiques collectifs.

695.760
[v.1]
Cop 1

Dissertation présentée pour
l'obtention du titre de
Docteur en Sciences Pédagogiques

par

Bernard LEBRUN

Directeur : Professeur L. VANDEVELDE

OCTOBRE 1977

2891995

A partir d'une communication de Jean CARDINET sur la typologie psychométrique des tests, nos recherches prirent naissance dans de longues et passionnées discussions entre le Professeur VANDEVELDE, Baghdad LAKHDAR et moi-même.

Ces recherches devaient s'épanouir dans le climat propice du Service de Didactique expérimentale de l'ULB, et aboutir aux développements relatifs à l'Evaluation Informative.

Nos remerciements iront d'abord au Professeur Louis VANDEVELDE, pour ses conseils judicieux, à Jacqueline HALLEUX pour ses encouragements, aux secrétaires Mesdames LEFEVRE et VAN GEYT, et à tous les membres du Service de Didactique expérimentale pour leur compréhension et pour leur dévouement.

Nos travaux eussent été incomplets s'ils n'avaient pas été fertilisés par les apports des enseignants, et nous retiendrons tout spécialement l'équipe de MARCHE-EN-FAMENNE, pour sa participation à la collecte des données nécessaires à la mise au point de nos méthodes.

Nous remercions également toutes les personnes qui, de près ou de loin, nous ont apporté leur aide dans cette entreprise de longue haleine, en particulier les membres des différents services de la Faculté des Sciences Psychologiques et Pédagogiques de l'ULB, Maurice HERION pour ses traductions et Raoul JAMBE pour ses corrections.

C'est enfin à mon épouse et à mes enfants que je dédie ce travail qui les a si souvent privés de ma présence.

EXPOSE DE LA THESE

HYPOTHESES FONDAMENTALES

L'évaluation pédagogique relève les transformations comportementales intervenues à la suite d'activités éducatives.

Dans la pratique courante, nous constatons que les enseignants concentrent leur intérêt sur les changements opérés chez l'INDIVIDU éduqué. Ils limitent leurs activités d'évaluation à la collecte de notes normées et à la communication de sanctions aux élèves.

Cependant, les conséquences COLLECTIVES de l'action éducative devraient aussi être évaluées, ce qui conduirait à la critique des moyens déployés pour atteindre les objectifs et attirerait l'attention sur la maîtrise des instruments de mesure utilisés.

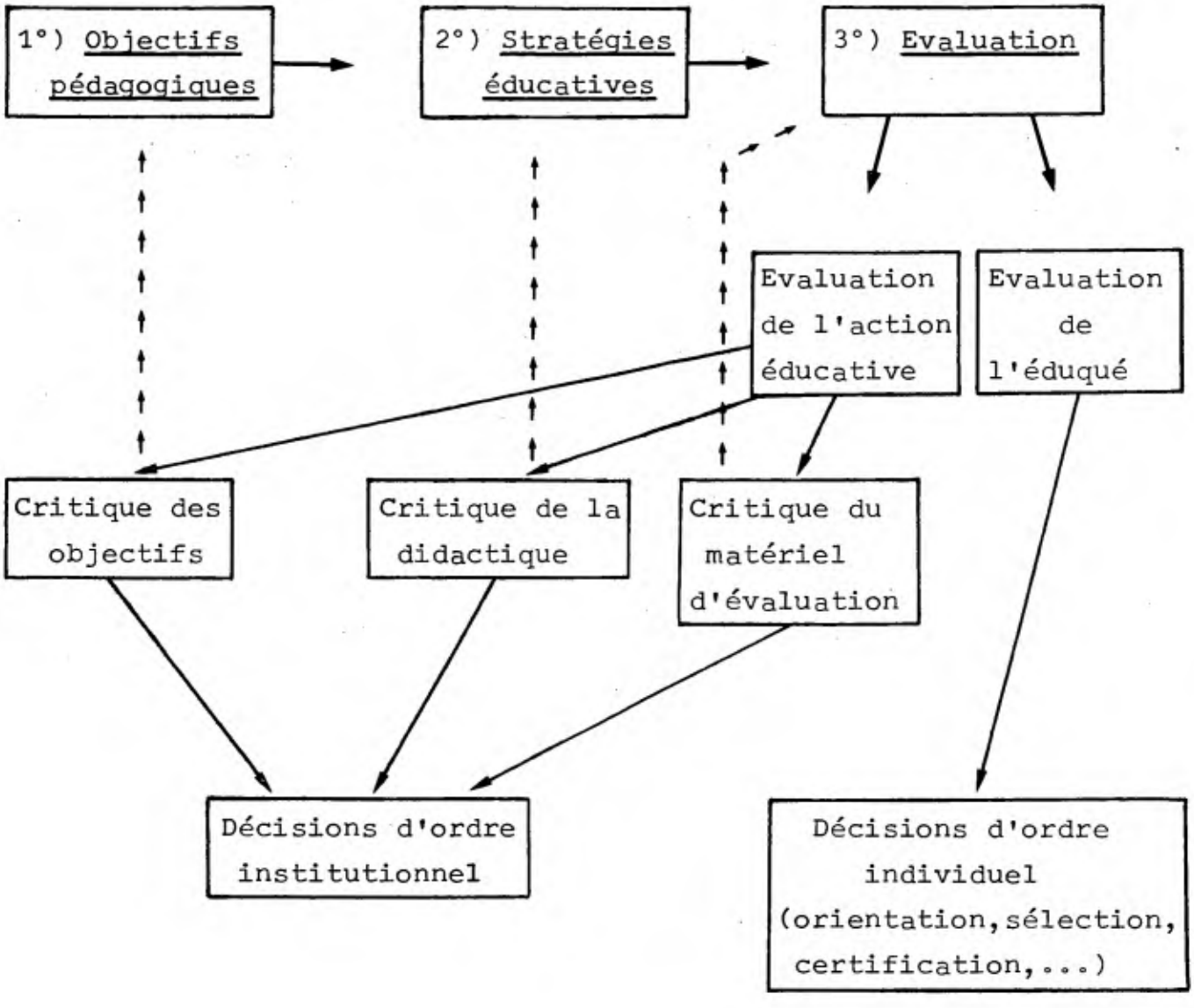
Nous nous trouvons ainsi devant l'alternative suivante:

- d'une part, une évaluation centrée sur l'éduqué, qui prend aussi bien en charge les effets de l'activité pédagogique que les apports individuels extérieurs,
- d'autre part, une évaluation destinée à mettre en évidence d'une manière privilégiée les effets de l'action éducative institutionnelle.

La nécessité s'impose de distinguer nettement ce qui est attribuable à l'institution scolaire des autres influences éducatives. La mesure de la pertinence de l'action pédagogique se révèle par l'estimation de résultats collectifs, puisqu'il n'est pas possible de déterminer, dans les résultats individuels, la part imputable à l'école et celle qui est propre à l'élève.

Schématiquement, nous situons l'évaluation, dans le processus pédagogique, de la manière suivante :

473173



Nous estimons qu'il est possible de proposer aux éducateurs des méthodes efficaces et adéquates pour évaluer les résultats de leur action au travers des effets collectifs qu'elle a engendrés. Nous donnons à l'ensemble de ces méthodes le nom d'EVALUATION INFORMATIVE.

Ce type d'évaluation de l'action éducative utilise des outils non sophistiqués et ne s'oppose nullement aux procédés statistiques classiques.

D'une part, ces méthodes sont commodes, c'est-à-dire compréhensibles pour l'utilisateur dépourvu de formation statistique poussée (= accessibilité technique).

D'autre part, ces méthodes sont maniables, c'est-à-dire simples et rapides, ne nécessitant le recours à des auxiliaires extérieurs (ordinateur,...) qu'en fonction des exigences et des possibilités des utilisateurs (= accessibilité matérielle).

Au cours de nos recherches, nous avons fait un certain nombre de constatations , qui ne ressortissent pas à notre objectif principal, mais qui sont de nature à nuancer le champ d'application des méthodes d'évaluation informative. Ces observations ont porté sur les changements d'attitude opérés chez les éducateurs à la suite de nos interventions.

1. Les informations recueillies sont exploitables pour une communication aux bénéficiaires de l'enseignement (bulletin,...), mais elles servent surtout aux prises de décision, en connexion avec les informations provenant d'autres sources.

2. Les enseignants sont amenés à modifier leur action éducative :
   - en définissant avec précision leurs objectifs;
   - en adaptant leurs interventions didactiques;
   - en soignant la conception et la réalisation de leur matériel d'évaluation.

3. Les responsables de l'enseignement changent certaines modalités du processus éducatif (objectifs, programmes, structures, formation des éducateurs,...).

4. Nos méthodes se prêtent aussi bien à une évaluation de portée limitée, mais aux effets immédiats (évaluation formative), qu'à une évaluation en fin de curriculum sur base expérimentale (évaluation sommative).

INTRODUCTION : Situation de l'évaluation dans l'action pédagogique

Mise au point terminologique

Nous aurions souhaité que le terme "évaluation" ne soit pas affecté à la fois à l'ensemble d'un processus et à la décision qui en est l'aboutissement.

Il nous semble indispensable de distinguer nettement le processus d'évaluation, qui recouvre l'ensemble des démarches entreprises, de la conclusion décisive de ce processus.

Nous continuerons cependant à appeler "évaluation" l'ensemble du processus, comme toutes les sources d'information qui entretiennent cette regrettable confusion, mais nous réserverons le terme "décision" à sa phase finale. Nous espérons que les praticiens, à qui nos travaux sont avant tout destinés, y trouveront leur compte.

Schéma général de l'action éducative

En suscitant, modifiant ou freinant certains comportements cognitifs, affectifs et psychomoteurs des élèves, l'éducateur adopte, en principe, une stratégie propre à rencontrer les options de l'institution qui l'a mandaté.

L'action pédagogique comprend trois étapes :

1. la définition des objectifs, à long, à moyen et à court termes;
2. l'utilisation de stratégies didactiques en vue d'atteindre ces objectifs;
3. l'évaluation des résultats enregistrés, afin de remédier à des carences observées ou d'établir un bilan de l'action engagée.

1. LA DEFINITION DES OBJECTIFS

Les concepts relatifs aux objectifs de l'éducation sont souvent vagues et imprécis, surtout si ces objectifs ont été récemment introduits, ou valorisés, par les courants pédagogiques.

" Il s'agit de développer la personnalité de chaque élève et de s'efforcer de le conduire le plus loin possible, dans le sens d'une qualification professionnelle qui corresponde à sa nature, à ses goûts et à ses possibilités. ( ) L'école doit constamment miser sur le changement et le pouvoir de changement de l'individu. Elle doit se tourner résolument vers l'avenir et s'encombrer le moins possible d'accumuler, sous forme de points, des traces du passé." (Circulaire ministérielle du 6/10/1969, reproduite dans "Faire le point sur ... l'enseignement secondaire rénové." (1974)

Dans un article récent, G. DE LANDSHEERE (1) reprend différentes propositions émises par de nombreux auteurs. De sa tentative de clarification, nous retiendrons trois grands niveaux d'objectifs :

a) Au sommet de la hiérarchie figurent les <u>finalités</u> de l'éducation, parfois appelées à tort objectifs généraux (ex.: libérer l'homme de toutes les aliénations). Ces grandes options, déterminées par les responsables politiques de l'éducation, ne semblent pas évaluables par les méthodes que nous proposons.

b) DE LANDSHEERE oppose les "<u>objectifs de maîtrise</u>", qui ne seront atteints que par des comportements totalement déterminés (ex.: la table de multiplication des dix premiers nombres), aux "<u>objectifs de transfert</u>", qui tendent à rendre l'étudiant capable de transférer les comportements acquis à des situations nouvelles. Comme il est impossible de prévoir toutes les difficultés susceptibles de se poser au cours d'une existence, l'éducateur sera contraint, dans ce dernier cas, de se limiter à des échantillons qu'il espère représentatifs.

c) Enfin, les "<u>objectifs médiateurs</u>", "<u>intermédiaires</u>" ou encore "<u>opérationnels</u>", constituent de simples étapes vers le terme d'un apprentissage (ex.: apprendre la méthode de la division écrite). Ils sont atteints par des initiatives concrètes prises au sein de la classe et bien délimitées dans le temps.

---

(1) G. DE LANDSHEERE - Pour un lexique des objectifs de l'éducation. PAEDAGOGICA EUROPAEA - Vol. XI - 1976 n°1 pp 31 - 35

Si ces derniers objectifs sont facilement reconnus par les éducateurs, il n'en va pas de même pour les objectifs de maîtrise et de transfert, qui sont souvent confondus.

1. Objectifs de maîtrise

L'instruction développe certains comportements privilégiés : l'enregistrement, la redécouverte, les activités répétitives, l'utilisation de l'acquis et la synthèse de cet acquis (1). L'enseignement porte sur des objets imposés par le système culturel. Il tend par conséquent à provoquer la convergence des comportements. L'activité didactique se voit essentiellement orientée vers la transmission d'informations, par le canal de matières et de programmes. Elle conduit à l'enculturation de l'individu.

2. Objectifs de transfert

L'éducation prétend aussi viser l'émancipation de l'individu, tant à l'égard des contraintes culturelles que du système éducatif lui-même. Elle doit alors se détacher de l'objet sur lequel portent les exercices intellectuels pour se focaliser sur les opérations. Cette formation tendrait à faire acquérir des aptitudes et des attitudes qui ne s'épanouiraient pas - ou peu probablement - sans entraînement systématique. Elle conduirait l'individu à l'autonomie.

A chaque composante intellectuelle de la formation correspond une opération mentale - compréhension, analyse, application, production, jugement (2) - et à chaque composante affective, on peut associer une attitude - recevoir, répondre, valoriser (3) - susceptible d'influencer l'éducation des aptitudes.

2. LA MISE EN OEUVRE D'ACTIVITES DIDACTIQUES

Pour atteindre les objectifs de transfert, les stratégies suggérées aux éducateurs sont souvent vagues et mal définies.

(1) VANDEVELDE L. et GUISLAIN G. Lexique relatif au cours de pédagogie psychologique et expérimentale. Presses Universitaires de l'ULB - 2e édition - 1977 - 39 p.

(2) VANDER ELST P. Prolégomènes à une pédagogie possible. Thèse en cours de publication. Université Libre de Bruxelles. 1976

(3) JAMBE et MASAI-PERL - L'éducation de l'affectivité - LABOR 1976

On parle, par exemple, de travail en équipe, de non-directivité... Or, si paradoxal que cette constatation paraisse, on ne conduira vers l'autonomie qu'en mettant en place des processus éducatifs contraignants, systématiques et souvent individuels.

Les recherches de FRYDMAN, dont nous reparlerons dans l'annexe 5, montrent que l'aptitude à utiliser des ouvrages de référence (dictionnaire, encyclopédie,...) est éducable dès l'enseignement primaire. En outre, cette compétence, à partir du moment où elle est acquise, s'accompagne de transfert au niveau des aptitudes mentales. Pour y arriver, il s'avère cependant nécessaire de soumettre les élèves à un entraînement systématique, dans le cadre duquel le rôle assumé par l'enseignant est un facteur déterminant.

Si l'enseignement a pour objectif la maîtrise de connaissances, il se fonde justement sur des exposés, dont les qualités essentielles sont la clarté et l'accessibilité. Le mode d'enseignement le plus prisé actuellement vise à la redécouverte des notions par les élèves et l'instrument le plus utilisé est la question. Dans une leçon de type classique, la question est au service de l'information que les élèves reçoivent. Elle sert surtout d'ajustement du niveau de la communication aux besoins du groupe. Cette méthode risque de confiner toute la stratégie didactique dans un va-et-vient continuel de questions et de réponses entre l'enseignant et les enseignés. A la limite, l'enseignant se livrerait à une maïeutique stérile, qui acculerait l'élève à donner la réponse attendue, même si l'énoncé de la question n'est pas correctement compris...

* * * *

Qu'il vise des objectifs de maîtrise ou de transfert, ou même des objectifs opérationnels, l'éducateur recourt à ce que nous appelons des "DECLENCHEURS". D'une manière très générale, le terme "déclencheur" désigne toute stratégie didactique destinée à réunir les conditions affectives et intellectuelles susceptibles de provoquer des comportements déterminés.

Le contenu du déclencheur, c'est-à-dire l'objet qui supporte l'activité, se distingue de son mode de présentation, qui en détermine la forme. La gamme des instruments mis à la disposition des éducateurs s'étend chaque jour. Nous observons actuellement l'intrusion en force des moyens audio-visuels, qui modifient radicalement la manière d'aborder de nombreux problèmes (1).

Deux formes particulières de déclencheurs retiennent l'attention :

1. Les déclencheurs explicites se présentent sous des formes très diverses : invitations verbales, mise en présence de réalités ou de leurs représentations, situations de problèmes, questions,...

2. Les déclencheurs implicites n'expriment pas une invitation formelle à se livrer à l'activité souhaitée. Ils visent à susciter chez l'élève des comportements "spontanés". Certains niveaux d'attitudes (vouloir recevoir, satisfaction à recevoir, vouloir répondre, satisfaction à répondre) ne peuvent être observés qu'en l'absence de déclencheurs explicites.

Par ailleurs, ne sous-estimons pas les ressources qu'offrent à l'enseignant les déclencheurs involontaires, imprévus, fortuits. Leur exploitation est souvent à la base d'interventions pédagogiques fructueuses, mais dépend de la claivoyance et de la perspicacité de l'éducateur.

* * * *

Un point capital, sur lequel nous nous permettons d'insister, c'est la nécessité de ne pas confondre l'entraînement didactique avec l'évaluation. Evaluer n'est pas éduquer, même si des exercices d'évaluation contribuent d'une manière non négligeable à certains apprentissages.

C'est au niveau des objectifs de transfert que cette confusion se fait le plus sentir. "Par des questions et des exercices appropriés, évaluer et en même temps stimuler le développement de telle ou telle fonction intellectuelle." (Circulaire du Ministère de l'Education nationale du 2 avril 1976 sur l'Evaluation des études dans l'enseignement secondaire).

---

(1) HALLEUX J. et LEBRUN B. L'audio-visuel au service de l'éducation. Presses Universitaires de Bruxelles. 1977

Comme le remarquent judicieusement JAMBE et MASAI-PERL (1976 - p. 38) :"On a tenté d'estimer à quel niveau de compréhension, d'analyse ou de synthèse étaient arrivés les élèves, plutôt que de mettre en oeuvre une didactique réellement appropriée à l'acquisition de ces aptitudes. Nul ne s'étonnera que les constatations soient souvent négatives. On reproche volontiers aux individus leur incapacité de réaliser des performances qu'aucun entraînement n'a préparées."

Avec DE LANDSHEERE (1972), nous stigmatisons ces procédés trop fréquents :"Combien d'échecs scolaires ne sont-ils pas dus aux belles questions d'"intelligence", posées au terme d'un enseignement qui ne l'a pas cultivée?"

3. L'EVALUATION

L'évaluation est le "processus qui consiste à décrire, recueillir et fournir des informations utiles pour porter un jugement décisif en fonction de diverses possibilités" (STUFFLEBEAM et Coll. 1971).

Elle ne constitue donc pas une opération isolée, mais un processus complexe qu'il convient de cerner avec beaucoup de précautions.

Dans la lignée de SCRIVEN(1967), très écouté en matière d'évaluation, les auteurs anglo-saxons soulignent la différence entre "mesure" et "évaluation".

D'une part, la mesure strictement quantitative refuse tout jugement de valeur sur la performance. Ses qualités essentielles sont la précision et l'objectivité.

D'autre part, l'évaluation tend à porter un jugement sur la qualité des résultats enregistrés.

BLOOM (1970) nous fournit une définition très détaillée des différents concepts que recouvre la notion d'évaluation :

Le "MEASUREMENT" s'applique à la recherche de résultats quantitatifs que l'on met en rapport avec des notes. Elle est indispensable pour la classification, la prédiction et l'expérimentation.

La validité, l'objectivité et la fidélité constituent les impératifs majeurs de celui qui mesure. La source principale d'erreur réside dans l'environnement, c'est-à-dire dans les circonstances qui entourent la prise de mesure.
Deux problèmes s'imposent actuellement au chercheur :
1. l'additivité des données,
2. la recherche d'instruments de plus en plus précis.

L'"EVALUATION" consiste à appréhender des changements intervenus chez les étudiants. Les données qu'elle traite sont aussi bien qualitatives que quantitatives. Les circonstances l'influencent également beaucoup.
En pratique, l'évaluateur se borne à analyser les changements intervenus dans des circonstances bien précises, en évitant toute sommation des notes. Si l'évaluation conduit à la certification, au placement, à la promotion de l'individu, elle permet aussi de s'assurer de la valeur d'une méthode, d'un programme, d'un enseignant, d'une séquence d'enseignement (curriculum).
La mise en évidence des effets d'un traitement sur l'individu ne sert qu'à rendre explicites son évolution et les changements intervenus.
L'évaluation ainsi définie pose des problèmes variés et difficiles. Elle exige la parfaite maîtrise des instruments et des techniques qui fournissent les résultats; elle se concentre sur les objectifs fixés au préalable par l'éducateur et identifie les expériences et les circonstances responsables des changements significatifs. Cette conception se rapproche très fort de ce que nous appelons généralement l'évaluation formative.

L'"ASSESSMENT" étudie l'interaction individu-milieu en inventoriant les caractéristiques des élèves dans leurs relations avec un environnement particulier, une tâche ou une situation déterminées. Des moyens d'investigation multiples sont mis en oeuvre pour analyser les circonstances qui entourent le travail, l'étude, la vie de l'individu; pour isoler ses besoins, ses attitudes, ses aptitudes, les pressions exercées sur lui.

L'objectif est la définition d'un modèle (construct) ou l'élaboration d'une théorie expliquant le comportement. On utiliserait l'"assessment" pour la prédiction, la sélection, l'expérimentation et la classification.

BLOOM s'attache ici à mettre en évidence les buts généraux que poursuivent les évaluateurs. En reprenant ses idées fondamentales, nous avons abouti à une approche chronologique des moments qui jalonnent le processus d'évaluation.

1°) LA PRISE DE MESURE

Cette première phase est marquée par l'intention de l'évaluateur d'entamer le processus d'évaluation. A l'aide de DECLENCHEURS appropriés, il suscite les activités qu'il veut évaluer. Sa préoccupation fondamentale consiste à obtenir que l'activité déterminée par un déclencheur se développe effectivement, c'est-à-dire sans donner lieu à des substitutions non désirées.

Pour s'assurer que l'élève s'est effectivement livré à l'activité souhaitée, l'évaluateur accompagne les déclencheurs de "REVELATEURS". Le plus souvent, ceux-ci se présentent sous la forme de questions, qui contraignent l'élève à se livrer à l'activité attendue. Donc, le révélateur est la pièce maîtresse de l'évaluation.

En fonction des objectifs éducatifs poursuivis, c'est de l'adaptation du révélateur à la nature de l'information recherchée que dépend la pertinence de la décision issue de l'évaluation. Toute modification du révélateur (choix multiple au lieu de question ouverte, par exemple) risque d'entraîner un type d'activité fort différent.

Le danger existe, cependant, de confiner l'évaluation dans la prise de mesure. C'est ce qui arrive lorsque les enseignants affectent leurs questions de valeurs absolues et décisives, sous la forme d'un capital de points accordés a priori. Il est même de bon ton, dans un souci de prétendue "objectivité", d'éviter tout retour en arrière sur le verdict rendu par les questions... Les pondérations prévues pour chacun des éléments constitutifs de

l'"examen" ou du "test" sont appliquées mécaniquement, et l'enseignant se contente d'enregistrer, d'additionner les résultats obtenus, qui consacrent, sans autre forme de procès, la réussite ou l'échec de l'élève.

STUFFLEBEAM et Coll. (1973) condamnent vigoureusement cette conception trop étroite et trop mécanique de l'évaluation, qui ne se justifie guère que dans les sciences d'observation où la mesure des phénomènes est essentielle. La limitation de l'évaluation à la prise de mesure contredit la dynamique de l'éducation qui se doit de développer une action concertée visant à un changement de comportement. Quand le météorologiste saura modifier les phénomènes naturels au lieu de se borner à les mesurer ou à les prévoir, il sera confronté à des difficultés que nous pouvons aisément imaginer : comment concilier les intérêts de chacun? Toute catastrophe lui sera imputée. De même, l'ambition de l'éducateur de modifier l'ordre naturel, livré autrement au hasard, implique de sa part une acceptation de responsabilités qu'il est vain d'éluder.

Les circonstances qui entourent la prise de mesure ne manquent pas d'influencer fortement les résultats obtenus. Nos observations ont montré que le matériel de mesure est très sensible (composition, longueur, forme, gradation des difficultés, lisibilité,...). De fortes modifications apparaissent dans les réponses lorsqu'un révèlateur se présente sous forme de choix-multiple plutôt que de réponse ouverte. Même des changements mineurs de formulation provoquent parfois des bouleversements spectaculaires. La position d'un révèlateur dans une épreuve a également une grande importance, ainsi que la longueur du travail.

Les effets dus aux individus testés sont nombreux. Relevons notamment :

- le degré de préparation de l'étudiant au type de travail à exécuter (readiness), sa familiarité avec la forme et le contenu du révèlateur (stratégie à l'égard du choix-multiple, par exemple). Beaucoup d'élèves sont surpris par l'originalité de certains révèlateurs, surtout lorsque ceux-ci visent des opérations peu fréquentes dans la pratique scolaire.

- les sanctions liées à l'évaluation, et leur présentation aux élèves. L'attitude des élèves varie fortement selon qu'ils sont conscients qu'une épreuve conduit à des décisions importantes ou ne constitue qu'un test de routine à l'impact négligeable.
- la transmission des résultats de l'évaluation. SKINNER (1968) souligne l'influence favorable d'une connaissance rapide des résultats.
- la disponiblilité, c'est-à-dire le temps et l'énergie qu'un élève accepte de consacrer à une tâche. La trop grande longueur d'une épreuve, ou le temps limite qui lui est imparti, constituent des facteurs parasites qu'il importe de neutraliser.

2°) L'ANALYSE DES DONNEES

a) L'évaluateur se trouve en présence d'un ensemble de réponses fournies par les élèves aux déclencheurs et aux révèlateurs. Dans la plupart des cas, les réponses ne sont pas directement exploitables : une transformation doit les rendre accessibles aux traitements envisagés. Une même observation se prête à une ou à plusieurs appréciations selon l'objet visé et la sévérité des exigences. Ainsi, en langue maternelle, on peut tenir compte séparément ou cumulativement de l'orthographe, de la correction de la formulation, de la précision des idées,... mais on peut aussi fixer plusieurs critères de niveau (réponse complète et correcte, tolérance de certaines erreurs,etc.). Tout dépend du but poursuivi par l'évaluateur.

Chaque observation, enregistrée selon un critère défini, constitue une unité d'analyse que nous appelons "item". Une même question fournit éventuellement plusieurs "items".

L'item (en anglais : élément) est donc défini :
- par la nature de l'information prélevée;
- par les critères qui sont associés à l'item.

Dans notre conception de l'évaluation centrée sur l'étude des effets collectifs du processus éducatif, l'item donne lieu à une note dichotomique : le comportement attendu se manifeste ou ne se manifeste pas.

b) Le traitement quantitatif des items exige des techniques adaptées aux buts assignés à l'évaluation, de manière à dégager, des données initiales disparates et souvent peu explicites par elles-mêmes, un maximum d'informations interprétables. En fonction de la décision à prendre à la fin du processus d'évaluation, le traitement sera orienté tantôt vers un classement, tantôt vers un niveau de maîtrise, tantôt vers un diagnostic.

La rigueur exige qu'un traitement des données réponde à trois conditions, dont nous empruntons l'essentiel à STUFFLEBEAM et Coll. (1971) :

1. Ces auteurs appellent "unité d'analyse" le niveau auquel se font simultanément les comparaisons, les associations et les rapports. L'unité d'analyse - qui correspond à notre définition de l'item - dépend du niveau de la prise de décision, de l'information que celle-ci nécessite, du degré d'organisation des données de base, et, dans les dispositifs expérimentaux, de la méthode employée pour l'échantillonnage.

2. La méthode d'analyse, c'est-à-dire la technique de traitement des items, dépend de l'unité d'analyse, de la quantité d'information nécessaire et du niveau auquel l'évaluateur travaille.

3. Dans la présentation de la situation, on tient compte des besoins en équipement, non seulement de l'infrastructure (ordinateur, programmes, personnel,...), mais aussi de la quantité de données disponibles.

Ayant choisi une méthode d'analyse, l'évaluateur soumet ses données de base à l'application des instruments. Il importe toutefois, conformément aux hypothèses fondamentales que nous avons formulées, que cette application soit parfaitement maîtrisée, et par conséquent que les outils soient communicables et accessibles.

DAGNELIE (1975 - p. 18) prétend que "l'utilisateur doit avoir une idée suffisamment précise des principes généraux et des conditions d'application de ces méthodes (d'analyse), mais il doit pouvoir consacrer essentiellement son attention à l'interprétation des résultats obtenus." Loin de nous l'idée de minimiser l'importance de l'interprétation. Il nous semble cependant indispensable

que celle-ci se fonde, non pas sur une appréhension vague, superficielle des traitements, mais bien sur leur maîtrise, par une perception approfondie de toutes les implications de la technique et de toutes les hypothèses qui la sous-tendent. Ainsi, comme le précise MATALON (1965 - p. 17), "il est illusoire de chercher à "mesurer" une variable quelconque si on ne sait rien de l'instrument employé. On pourra dire qu'un sujet est supérieur à un autre, par exemple, s'il a résolu plus de problèmes. Mais cela suppose qu'on a déjà vérifié (ou admis) que ces problèmes sont, en un certain sens, équivalents." L'additivité des données constitue, à notre avis, une des conditions les plus astreignantes dans un traitement valable de données pédagogiques.

c) Interpréter les résultats, c'est adapter les informations fournies par la technique à chacune des situations particulières. Tout en respectant le contenu initial des données, l'évaluateur, pour rendre ses observations significatives, rassemble et organise les informations susceptibles d'apporter plus de certitude à la décision.

L'interprétation s'effectue sur la base des éléments suivants :

1) les objectifs éducatifs définis;
2) les moyens didactiques mis en oeuvre pour atteindre ces objectifs;
3) la forme des déclencheurs et des révélateurs, ainsi que les intentions qui ont présidé à leur élaboration;
4) les circonstances qui ont entouré la prise de mesure. Ce dernier facteur, trop souvent méconnu, nous paraît essentiel. L'influence des circonstances, que nous appelons "effet de contexte", est négligée par la plupart des constructeurs d'épreuves, qui omettent de renseigner les utilisateurs sur les conditions à remplir pour que les résultats soient comparables aux normes proposées. Il est, notamment, délicat de comparer les performances à une même épreuve présentée en prétest et en post-test aux mêmes élèves, puisque le contexte a varié d'une manière difficilement appréciable.

Certaines circonstances sont fortuites. Elles s'imposent en quelque sorte à l'évaluateur, qui devra en tenir compte a posteriori. Par contre, de nombreuses circonstances sont contrôlables, parce qu'elles sont prévisibles. Il est donc possible d'en déterminer l'impact probable avant d'entamer la prise de mesure.

3°) LA DECISION

En matière d'éducation, les analyses les plus rigoureuses et les précautions les plus attentives ne peuvent éliminer complètement le caractère aléatoire de toute décision.

Nous comprenons volontiers le désarroi de tous ceux qui, à quelque niveau qu'ils se situent dans l'enseignement, sont amenés à décider. Nous n'avons pas de solution définitive à leur proposer, mais nous pensons qu'une meilleure compréhension du mécanisme de l'évaluation et la maîtrise des techniques que celle-ci met en oeuvre les aideront à assumer la démarche inéluctable de la décision. Nous espérons leur rendre ainsi la sérénité qu'une approche parcellaire et incomplète du problème leur avait fait perdre.

Nombreux sont les auteurs qui s'intéressent, à l'heure actuelle, aux mécanismes que nous évoquons. STUFFLEBEAM et Coll.(1971) reconnaissent quatre grandes étapes à une prise de décision dans un contexte éducatif :

- AWARENESS
- DESIGN
- CHOICE
- ACTION

Ces étapes méritent d'être brièvement développées et assorties de considérations relatives aux situations d'évaluation que nous envisageons dans ce travail.

1° Prendre conscience qu'une décision est nécessaire (AWARENESS)

L'éducateur effectue trop souvent des évaluations mécaniques, notamment pour répondre aux exigences institutionnelles (bulletins périodiques, nombre minimum de travaux à effectuer par les élèves,etc.).

D'une manière générale, la prise de conscience est liée à deux situations éducatives :

a) l'hypothèse de l'existence d'une divergence entre les objectifs poursuivis par l'éducateur et les effets réellement obtenus;

b) les échéances prévues par l'institution (examens, diplômes, orientation, sélection,etc.).

2° Etablir un plan de la situation de décision (DESIGN)

Ce plan fait intervenir :

a) l'inventaire des décisions possibles et des alternatives éventuelles;

b) la ou les personnes habilitées à décider;

c) la détermination des critères fondamentaux;

d) la fixation des règles qui seront appliquées; certaines sont imposées par l'institution, beaucoup d'autres laissées à l'appréciation des évaluateurs;

e) les contingences temporelles relatives à la prise de décision et les formes à respecter.

3° Choisir entre les différentes possibilités (CHOICE)

C'est la décision proprement dite. Il convient d'écarter, parmi les alternatives possibles, celles qui ne reposent pas sur des informations évaluables. Ici se pose évidemment le problème de la collecte d'informations valides et variées. Il arrivera souvent que la seule décision prise prévoie la recherche de nouvelles informations... Enfin, avant d'appliquer les règles fixées dans le plan, il faut vérifier leur adéquation à la situation évaluée.

4° Veiller à donner une suite aux décisions prises (ACTION)

La décision doit encore être appliquée : son utilité se mesure à ce prix. les personnes chargées de l'exécution sont désignées. Elles devront, à un moment donné, rendre compte des modalités d'application.

Quelle est la situation de l'évaluateur dans la prise de décision? "Faire un choix ressort de la responsabilité du décideur. La tâche essentielle de l'évaluateur est de fournir de l'information pour aider le décideur à faire son choix." (STUFFLEBEAM et Coll. - 1971 - p.107). Dans l'enseignement toutefois, le décideur et l'évaluateur ne sont le plus souvent qu'une seule et même personne. L'enseignant est donc à la fois juge et partie! Il doit assumer les risques de cette situation équivoque - où l'on voit parfois des élèves affublés d'étiquettes définitives dans des dossiers qui aboutissent on ne sait trop où - en attendant de pouvoir compter, comme le suggèrent certains auteurs, sur l'intervention d'évaluateurs spécialisés et indépendants.

4°) LA COMMUNICATION

Il n'entre pas dans nos intentions de nous étendre longuement sur la communication de la décision, mais les éducateurs se doivent de lui accorder le plus grand soin.

Nous trouvons dans des directives récentes (1) destinées aux enseignants des indications intéressantes sur les propriétés de cette communication :

"(Les notations) sont d'autant plus valables qu'elles sont significatives pour ceux à qui elles sont destinées, c'est-à-dire aux élèves et aux parents.

D'une manière générale, il convient de distinguer la note chiffrée encore appelée cote ou score, l'appréciation globale exprimée par niveau et l'appréciation rédigée.

Suivant sa nature, l'épreuve peut être évaluée par compte ou décompte de points; on parle alors de cotes sur 5, 10, ... 100.

Par contre, si l'on souhaite situer l'élève dans un groupe: classe, année d'études dans l'établissement ou même dans plusieurs établissements, l'expérience et la recherche universitaire recommandent, parmi d'autres techniques, l'utilisation d'une échelle

---

(1) Ministère de l'Education nationale et de la Culture française - Circulaire du 2 avril 1976 intitulée :"Evaluation des études dans l'enseignement secondaire."

d'appréciation à cinq niveaux, s'exprimant par notations littérales TB, B, S, F, I ou chiffrées 1, 2, 3, 4, 5, et qui convient bien pour l'évaluation d'épreuves complexes ou, plus généralement, pour des évaluations globales.

L'appréciation rédigée permet toutes les nuances de l'expression écrite et communique le résultat d'une analyse plus fine de l'épreuve."

Ces directives appellent quelques commentaires :

1. Les intentions de départ sont excellentes : elles soulignent la nécessité d'expliciter les significations que les notes impliquent.
2. La communication est limitée à deux types de bénéficiaires : les élèves et les parents. Elle ne s'adresse ni aux autres éducateurs (Conseil de classe), ni aux responsables de l'enseignement.
3. En cours de processus, l'utilisation systématique de notes normées, qu'elles soient quantitatives ou qualitatives, éclaire mal les décisions éducatives, même si on tempère ces notes par l'adjonction de commentaires verbaux plus nuancés. Si l'expérience montre effectivement que la notation spontanée des enseignants est de type normatif, nous craignons que la recherche universitaire ne puisse cautionner une prise de position aussi catégorique!

La communication est pourtant nécessaire, ne fût-ce que pour renforcer l'activité des élèves, mais un tel objectif s'accommode mal de la brutale constatation chiffrée ou littérale résultant d'une analyse sommaire. Il convient donc, après avoir tenu compte des effets imputables à l'action éducative elle-même, d'adapter la communication aux circonstances, sans craindre la "subjectivité", pour provoquer chez l'élève l'attitude attendue.

On trouvera ci-après un résumé schématique des différentes étapes que nous avons développées dans ce chapitre.

Schéma général de l'action éducative

I DEFINITION DES OBJECTIFS

a) Finalités
b) Objectifs de maîtrise et objectifs de transfert
c) Objectifs médiateurs, intermédiaires ou opérationnels

II MISE EN OEUVRE D'ACTIVITES DIDACTIQUES

Présentations de déclencheurs | explicites
| implicites

Exploitation de déclencheurs involontaires

III EVALUATION

1. Prise de mesure
   - Choix de déclencheurs appropriés
   - Construction de révélateurs spécifiques
   - Prévision des circonstances entourant la prise de mesure
2. Analyse des données
   - Transformation des réponses en unités d'analyse : les items
   - Traitement quantitatif des items
   - Interprétation des résultats
3. Décision
   - Prise de conscience qu'une décision est nécessaire
   - Etablissement d'un plan de la situation de décision
   - Choix entre les différentes possibilités
   - Exécution des décisions prises
4. Communication
   - en fonction des destinataires

CHAPITRE I FONDEMENTS THEORIQUES DE NOTRE PROPOSITION D'EVALUATION INFORMATIVE.

C'est dans la "typologie psychométrique des tests" de Jean CARDINET (1973) que nous avons trouvé l'instigation de nos recherches. Par la suite, la similitude existant entre la matrice hiérarchique de CARDINET et le scallogramme de GUTTMAN (1944) nous a conduit à étudier les développements donnés à ce dernier modèle, même si l'objet qui l'avait suscité est souvent fort éloigné de nos préoccupations.

1°) La typologie psychométrique des tests (1)

Cette typologie constitue une proposition très constructive pour enrichir la gamme d'instruments dont les organismes chargés de l'évaluation ont besoin. Elle se dégage de l'analyse des données contenues dans des tableaux à double entrée, où les lignes représentent les élèves et les colonnes les items.

Les données s'expriment sous la forme 1 (réussite) ou 0 (échec). L'auteur suggère de recourir au modèle additif de l'analyse de la variance pour dégager l'existence d'effets dominants.

La somme des carrés des écarts à la moyenne de l'ensemble de la matrice d'observations peut se décomposer en :

- une fraction due aux différences de moyennes des lignes,
- une deuxième fraction due aux différences de moyennes des colonnes,
- une troisième due à l'interaction lignes-colonnes,
- une part provenant de facteurs incontrôlés régis par le hasard.

La démarche consiste à maximiser tantôt la fraction de la somme des carrés totale due aux différences entre lignes (élèves), tantôt la part de la somme des carrés totale imputable aux colonnes (items).

---

(1) CARDINET Jean L'adaptation des tests aux finalités de l'évaluation. Revue de l'Ecole nouvelle n° 2-3 - 1973 DIDIER PARIS.

Quatre modèles fournis par CARDINET nous semblent mériter une attention particulière dans le cadre du processus d'évaluation pédagogique.

1. Le test de classement

D'après CARDINET, il est caractérisé par l'importance dominante des différences entre sujets. La matrice type qui le représente apparaît comme suit :

1	1	1	1	4
1	1	1	1	4
0	0	0	0	0
0	0	0	0	0
2	2	2	2	8

Pourcentage de réussite des élèves : 100 % ou 0 %.
Moyenne générale : 50 %.
Pourcentage de réussite des items : 50 %

La somme des carrés des écarts à la moyenne est entièrement imputable aux différences entre élèves. Les parts correspondant aux différences entre items et à l'interaction élèves-items sont nulles.
Ce test, différenciant parfaitement les compétences des élèves, conviendrait pour procéder à la sélection, à la promotion, etc.

Nous avons substitué le terme "différenciation" à celui de "classement" pour qualifier cette épreuve. L'ensemble à classer n'offre que la possibilité d'une distinction entre deux collections d'éléments nettement distincts et homogènes. Certains élèves dominent parfaitement le contenu des items, les autres absolument pas.

Si elle constituait l'objectif de l'évaluation, une sélection basée sur ce test serait vraiment efficace. Il conviendrait donc d'en contrôler parfaitement les fondements, c'est-à-dire le contenu des activités dont les résultats ont conduit à la différenciation.

2. Le test d'enquête

Pour CARDINET, il répond à une matrice définie par la maximisation des différences entre items.

1	1	0	0	2
1	1	0	0	2
1	1	0	0	2
1	1	0	0	2
4	4	0	0	8

Pourcentage de réussite des élèves : 50 %
Moyenne générale : 50 %
Pourcentage de réussite des items : 100 % et 0 %

Les professeurs, les responsables de l'enseignement trouveraient dans des épreuves de ce type de précieux renseignements sur le rendement des méthodes, d'adéquation des programmes, les succès ou les échecs selon les notions considérées.

Comme CARDINET, nous estimons que de telles épreuves mèneraient à l'évaluation du processus éducatif. Cependant, nous ne voyons pas la nécessité de joindre systématiquement les items à 100 % de réussite (supérieurs) aux items à zéro % de réussite (inférieurs). CARDINET avait le souci, pour présenter une typologie reposant sur un modèle exhaustif, de mettre en parallèle des épreuves caractérisées par une différenciation sur les marginales. Les épreuves de classement et d'enquête correspondent à ce souhait.
Dans des études longitudinales, une telle représentation donne effectivement une idée des marges entre lesquelles se situent les variations. Par contre, dans la pratique pédagogique, la grande différence entre les pourcentages de réussite aux deux groupes d'items ne constitue nullement l'élément essentiel de l'observation. C'est le caractère "extrême" de chacune des deux "familles" qui retient avant tout l'attention. Dans bien des épreuves n'apparaissent pas d'items supérieurs ou inférieurs. L'interprétation des items n'en est nullement remise en cause.

3. Le test d'inclination

D'après CARDINET, les différences entre items et entre élèves sont nulles et n'interviennent pas dans la somme des carrés des écarts à la moyenne. C'est l'interaction qui est responsable de l'ensemble de la variation. De plus, cette seule source de variation est parfaitement cohérente, puisqu'il existe une opposition entre la réussite à une collection d'items et l'échec à une autre.

1	1	0	0	2
1	1	0	0	2
0	0	1	1	2
0	0	1	1	2
2	2	2	2	8

Pourcentage de réussite des élèves : 50 %

Moyenne générale : 50 %

Pourcentage de réussite des items : 50 %

Ce test serait l'outil de l'orientation, de l'adaptation de l'enseignement aux dispositions particulières des élèves. Sur un modèle factoriel, il mettrait en évidence les caractéristiques affectives des individus (motivations, intérêts) selon un processus polaire.

Nous avons préféré le terme "contraste", qui rend mieux compte de la situation théorique, à celui d'"inclination". Ce test souligne l'opposition interne entre items d'une part, entre élèves d'autre part, plus qu'il ne débouche sur une décision d'orientation comme son intitulé original le fait pressentir.
Une telle matrice de résultats n'a jamais été mise en évidence, jusqu'à présent, dans une épreuve scolaire. Tout au plus observons-nous parfois des items isolés contrastant avec des familles d'items. Il n'en reste pas moins qu'une telle épreuve devrait constituer la pierre angulaire de l'orientation.

4. Le test hiérarchique

Pour CARDINET, les compétences des élèves et les exigences relatives des items se répartissent équitablement la somme des écarts à la moyenne. On observe à la fois une hiérarchie des élèves et des items.

0	0	0	1	1
0	0	1	1	2
0	1	1	1	3
1	1	1	1	4
1	2	3	4	10

Cette épreuve rendrait parfaitement compte de l'évolution de l'apprentissage chez un même individu

Remarquons qu'il s'agit en réalité d'un test de classement, dans le sens courant du terme. La concordance est parfaite entre les résultats obtenus à chaque item et ceux obtenus à l'ensemble de l'épreuve.

Dans la présentation de sa typologie, CARDINET insistait lui-même sur le fait que les matrices théoriques qu'il avait définies sont évidemment fort éloignées de celles que fournissent les épreuves réelles. Il déclarait alors n'avoir pas encore de solution pratique à présenter. Depuis, ses recherches se sont orientées vers l'application de la Théorie de la Généralisibilité. Par l'étude des sources de variance analysables sur les différentes facettes qui produisent des variations, il quantifie les fluctuations responsables de la variance-erreur, qui proviennent essentiellement des lois de l'échantillonnage.

Notre principal souci a été de tout mettre en oeuvre pour rendre les modèles utilisables dans le cadre de la pédagogie quotidienne. Nous ne prétendons pas avoir épuisé les ressources de la typologie, mais les processus que nous avons élaborés permettent

d'obtenir des indicateurs, sinon précis, en tout cas utiles, pour une exploitation rapide des principes énoncés par CARDINET.

2°) Le scallogramme (1)

Dans un article datant déjà de plus de 30 ans et repris par FISHBEIN (1967), GUTTMAN a proposé une approche originale de la mesure des attitudes, à partir de la comparaison par paires des observations effectuées. Le tableau type, appelé scallogramme, présente l'aspect d'un "parallèlogramme".

1	1	1	0	0	0	3
0	1	1	1	0	0	3
0	0	1	1	1	0	3
0	0	0	1	1	1	3
1	2	3	3	2	1	12

On obtient cette représentation par un choix judicieux des attitudes que l'on suscite et après une permutation adéquate des lignes et des colonnes. Le scallogramme est évidemment très proche de la matrice hiérarchique de CARDINET : il suffit de séparer les trois premières colonnes des trois dernières pour obtenir deux hiérarchies contrastées! Les utilisateurs n'ont d'ailleurs pas hésité à sacrifier les colonnes contrastées, qui n'apportent pas d'information supplémentaire. (LOEVINGER - 1948, GREENE - 1954, NIE et Coll. - 1975)

GUTTMAN insiste sur la nécessité de disposer d'un matériel de mesure de qualité, le hasard étant bien sûr générateur d'interaction. Il suppose audacieusement, dans son schéma, qu'aucune question n'est affectée par l'effet du hasard!

L'originalité de cette présentation des résultats réside dans le fait que les hiérarchies sur les marginales proviennent de l'analyse des données beaucoup plus que de considérations a priori. GAGNÉ (1967 - p. 29) souligne l'intérêt pédagogique de cette procédure :

---

(1) GUTTMAN L. A basis for scaling qualitative data. American Sociological Review, 1944 - 9, 139 - 150.

" Une méthode assez stricte peut être employée pour tester l'adéquation d'une structure proposée pour un programme d'enseignement. Cette méthode consiste à concevoir et à administrer un test qui a été spécialement construit pour fournir une information de type "réussite-échec" sur chaque unité de connaissance au sein d'une hiérarchie absolue. Les résultats d'un tel test sont ensuite analysés pour mettre en évidence l'interdépendance séquentielle des unités."

GUTTMAN propose un indice pour calculer la déviation de la matrice observée par rapport à la matrice-type. L'indice de reproductibilité (Rep) s'obtient par l'application de la formule suivante :

$$Rep = 1 - \frac{\text{nombre d'erreurs}}{\text{nombre de réponses}}$$

GUTTMAN ne fournit cependant que peu d'informations sur la manière de procéder pour mettre un scallogramme en évidence. C'est probablement ce qui a nui au développement de sa technique, au profit de la méthode proposée par LIKERT (1932) centrée, elle, sur la validité et la fidélité des questions.

Pour notre part, nous avons estimé que le scallogramme constitue un élément essentiel dans la recherche d'instruments d'évaluation en pédagogie. Tous les chercheurs qui se sont penchés sur cette configuration particulière se sont heurtés à de grosses difficultés, non seulement au niveau mathématique, mais aussi pour trouver des données qui satisfassent aux exigences du modèle.

Notre objectif ne sera pas de susciter des activités qui, a priori, sont destinées à former une hiérarchie, mais bien d'identifier, dans une perspective pédagogique, les observations correspondant au modèle.

CHAPITRE II PRINCIPES DE BASE DE L'EVALUATION INFORMATIVE

1. CONTRAINTES IMPOSEES PAR NOS HYPOTHESES FONDAMENTALES

a) Choix d'une méthode valide

La plus élémentaire prudence engage l'évaluateur à vérifier d'abord si le matériel utilisé correspond aux ambitions de sa didactique et de son évaluation. Nous entamons donc l'analyse d'une matrice de données par l'étude de la dimension "ITEM", par opposition à l'évaluation traditionnelle qui ne se préoccupe que de la dimension "ELEVE".

Une méthode soucieuse de mettre en évidence les effets collectifs de l'action pédagogique conduit à minimiser les effets individuels. L'analyse prioritaire de la dimension "item" fournit à l'évaluateur les informations recherchées. En effet, le comportement des items rend compte des résultats atteints par l'éducateur, puisque les élèves sont tous censés avoir tiré bénéfice de l'action entreprise.

Les élèves, bénéficiaires au premier chef de l'action pédagogique, se présentent ainsi comme de simples témoins privilégiés. Si l'évaluateur enregistre leurs transformations comportementales, ce n'est pas pour relever leurs performances, mais pour mettre en évidence, parmi les effets collectifs observés, ceux qui sont imputables au système éducatif et qui consacrent ainsi sa pertinence.

Par contre, les méthodes d'évaluation appliquées d'habitude dans l'enseignement s'orientent exclusivement vers le classement des élèves : elles entérinent une sommation de notes disparates que l'éducateur ne contrôle guère. Les barêmes et les étalonnages qui en résultent ne donnent de la situation réelle qu'une perception floue et peu susceptible de mener à une action efficace.

La validité de l'évaluation informative a été mise à l'épreuve dans de nombreuses situations pédagogiques. Cependant, la démarche expérimentale consistant à comparer des groupes soumis à des traitements différents n'est pas applicable. Nous mettons en évidence, pour le destinataire, le caractère utile des informations, qui sont ainsi justifiées par leur utilité même.

b) Choix d'une méthode commode

Notre proposition est surtout destinée aux acteurs de l'éducation, sans être indifférente aux besoins des expérimentateurs.

Les conclusions intéressent directement les ENSEIGNANTS, qui participent activement au processus éducatif et sont confrontés à l'utilisation quotidienne de l'évaluation. Mais le personnel d'encadrement et les responsables de l'enseignement (directeurs d'école, inspecteurs, rédacteurs de programmes,etc.) y trouvent également des éléments essentiels pour l'adaptation du système, en général.

Il faut que ces destinataires comprennent les fondements de leurs interventions évaluatives. Puisqu'ils ne possèdent pas une formation statistique spécialisée, nous avons évité les démonstrations compliquées : nous nous sommes attaché à définir des modalités d'évaluation dont les principes fondamentaux apparaissent clairement à tous les stades de l'exécution :

1. Le professeur est invité à intégrer son évaluation dans la continuité du processus didactique. Pour adapter ses instruments de mesure aux circonstances, il les construit la plupart du temps lui-même. En vue d'un traitement rapide des données recueillies, nous lui proposons une technique simple, qui ne demande que la lecture d'une table et dont les calculs sont quasiment exclus. Nous mettons à sa disposition (annexe 3) une présentation très simplifiée de la technique d'évaluation informative. Elle lui permet de prendre des décisions rapides, voire immédiates,mais limitées au cadre de sa classe.

2. Le responsable de l'enseignement, quant à lui, étudie les effets produits à plus long terme par un ensemble d'interventions éducatives portant sur un grand nombre d'élèves. Il s'entoure de précautions particulières en matière d'échantillonnage. Il choisit avec soin ses instruments de mesure et les outils de traitement des données. Les techniques d'évaluation informative que nous lui suggérons sont plus fines et plus variées que celles du professeur (annexe 1). Cependant, ces propositions n'exigent aucune formation statistique particulière.

3. Le chercheur spécialisé ne se satisfait pas d'une approche aussi pragmatique. Il trouve dans nos méthodes des indices utiles à la poursuite de l'analyse par des techniques plus élaborées (annexes 2, 4 et 5).

c) Conclusion

Notre objectif est précis : mettre à la disposition de l'éducateur des moyens d'évaluation plus valides et plus commodes. Nous nous posons cependant la question de savoir si l'enseignement, tel qu'il est pratiqué à l'heure actuelle, se prête à une véritable évaluation, et par conséquent, si le malaise que ressentent les éducateurs à ce sujet ne prend pas ses racines dans la conception et la mise en pratique des activités éducatives elles-mêmes.

2. CARACTERISTIQUES GENERALES DE NOS METHODES D'ANALYSE

Nous avons vu que le processus d'évaluation se décompose en quatre étapes bien différenciées, nécessitant chacune un examen particulier :

a) la prise de mesure,
b) l'analyse des données,
c) la décision,
d) la communication.

Sans que nous nous désintéressions pour autant des autres étapes, nos travaux ont essentiellement porté sur un aspect particulier de l'analyse des données : le traitement des informations consécutives à la prise de mesure et en vue d'une décision.

Au départ, nous avions espéré trouver parmi les instruments statistiques classiques une aide efficace pour atteindre nos objectifs. Nous nous sommes donc tourné vers l'analyse de la variance, mais après quelques résultats prometteurs, la non-normalité de nos distributions nous a conduit dans une impasse. L'exploitation des coefficients de corrélation s'est également avérée infructueuse.

En fait, nous avons rencontré les mêmes problèmes que BENZECRI (1973), qui exprime sa déception non sans une pointe d'agressivité : " Statistique n'est pas probabilité. Sous le nom de statistique mathématique, des auteurs ont édifié une pompeuse discipline, riche en hypothèses qui ne sont jamais satisfaites dans la pratique."

Les modèles probabilistes peuvent-ils, dès lors, nous apporter les solutions escomptées ? Comme BENZECRI, nous devons encore déchanter :

" Lors même que l'existence d'un ordre des choses probabiliste est avérée, on ne dispose généralement pas d'un modèle probabiliste complet (modèle qui est à la base de la statistique bayesienne, et aussi de la théorie de l'information).

(...) Le modèle doit suivre les données, non l'inverse. Autre trait fâcheux des mathématiques appliquées aux sciences humaines : l'abondance des modèles, forgés a priori, puis confrontés aux données par ce qu'on appelle des tests. Et tantôt le "test" sert à justifier un modèle où il y a plus de paramètres à ajuster que l'on a déterminé de données. Tantôt au contraire il sert à rejeter sévèrement comme invalides les plus judicieuses remarques de l'expérimentateur.

(...) Il est impossible d'éliminer d'un modèle les conceptions a priori qui ont permis au chercheur d'en imaginer les formules, formules certes confrontées avec l'expérience, mais seulement pour en ajuster les paramètres aux mesures faites. Ainsi, le modèle permet au mieux de prédire, non de comprendre."

Pour résoudre ses problèmes, BENZECRI a développé la méthode dite "analyse des composantes" que DAGNELIE (1975) définit de la manière suivante (p. 15) :

" L'analyse des composantes intervient dans l'interprétation des relations existant entre une série de variables interdépendantes. Son but principal est de condenser l'essentiel des informations données par ces variables interdépendantes, observées directement, en un nombre plus restreint de variables fondamentales indépendantes, qu'on ne peut observer directement. Tel est le cas notamment en biométrie, lorsque l'on a mesuré diverses caractéristiques (poids, taille,etc.) d'un ensemble d'organismes vivants, et qu'on désire condenser ces observations en un petit nombre de caractéristiques fondamentales (croissance en hauteur ou élongation, croissance en largeur, etc.)."

Nous souscrivons aux principes qui sous-tendent la méthode d'analyse de BENZECRI. Elle ne répond cependant pas aux objectifs que nous avons assignés à notre recherche.

a) le concept de la distance, qui est à la base de l'analyse des composantes, ne pose pas de problème majeur en ce qui concerne sa compréhension. Mais pour l'exploiter, on recourt à des développements qui considèrent chaque variable (item) comme une dimension. Le point de départ de l'analyse est constitué par une matrice qui comprend autant de dimensions qu'il y a d'items.

L'analyse des composantes accorde a priori une validité à chaque item, ce que nous nous refusons à admettre. Ce n'est que lorsque plusieurs items présentent un comportement similaire - nous pourrions parler d'une typologie d'items - que nous leur concédons de l'intérêt. Nous devons à tout prix déceler rapidement les items isolés, aux caractéristiques indéfinissables, pour les écarter afin qu'ils ne contaminent pas les groupes d'items chargés en informations fiables. Nous nous sommes donc tourné vers le concept de la CONCORDANCE. Si nous observons une correspondance entre le comportement de deux ou plusieurs items, nous posons l'hypothèse

qu'il existe une relation entre eux. L'interprétation nous montre si cette relation est liée aux effets pédagogiques étudiés.

Au cours de nos travaux, nous avons ressenti la nécessité de nous en tenir à une approche unidimensionnelle des problèmes. Toute proposition plus complexe se heurte invariablement à un refus de la part de nos destinataires. A fortiori, l'interprétation des résultats n'est accessible qu'à ceux qui maîtrisent tous les arcanes de la technique.

b) Notre analyse est orientée, non pas vers un objet, mais bien en fonction d'un usage précis. Il n'est pas question de suivre toutes les pistes que l'analyse nous révèle. Nous ne retenons que celles qui constituent des informations utiles pour l'action pédagogique. Nos objectifs sont limités par les contraintes matérielles et temporelles de la pratique éducative.

c) Chaque fois que les circonstances l'ont permis, nous avons appliqué l'analyse des composantes de BENZECRI aux données que nous avons exploitées. Nous verrons dans les annexes 1, 2 et 4 dans quelle mesure ses résultats sont compatibles avec ceux de nos méthodes.

3. CHOIX D'UN MODE DE NOTATION

A la source même de l'évaluation, voici l'élève, confronté à une activité particulière. Puisque, dans l'évaluation informative, ce n'est pas sa performance qui est visée, mais bien l'effet que l'éducation a produit, nous devons nous assurer que la difficulté a été franchie. Ou le but est atteint, ou il ne l'est pas. La seule notation nécessaire est donc 1 ou 0.A quoi bon tergiverser ? L'apport de nuances constitue une fuite qui risque de masquer la nature des effets enregistrés.

Pour le traitement des données par les techniques d'évaluation informative, les items doivent mener à une notation alternative, parfois aussi appelée notation "binaire" ou notation "dichotomique".

Ce mode d'appréciation présente des avantages décisifs dans le cadre de l'analyse des données, ainsi que le souligne MATALON (1965 - p.20): "(Avec la notation dichotomique), les problèmes sont considérablement simplifiés, non seulement parce qu'on réduit ainsi le nombre de termes intervenant dans les expressions, mais aussi parce qu'il devient alors possible, le plus souvent, de faire abstraction de la relation d'ordre entre les deux réponses."

Lorsqu'il s'agit de mesurer l'efficacité d'une didactique, la notation dichotomique s'impose. Comme l'écrit GAGNÉ (1967 - p.37): " La mesure de base à obtenir, qui fournisse les moyens de contrôle, revient à déterminer si l'étudiant a acquis ou n'a pas acquis les aptitudes spécifiques relatives à l'unité mesurée. Il ne peut être question d'introduire des degrés de maîtrise dans cette sorte de mesure, ni de tenir compte du concept de difficulté. Dès que de telles mesures de base sont effectuées, on peut par la suite poser des questions à propos d'autres catégories de différences individuelles, et ces dernières varieront probablement en intensité."

Nous formulons une réserve à propos de ce que dit GAGNÉ. Dans notre conception de l'évaluation informative, nous ne vérifions pas si l'étudiant a acquis quelque chose, mais bien si le comportement attendu est présent ou non au moment de la prise de mesure.

Au cours de nos travaux, nous avons insisté auprès des constructeurs d'épreuves pour que leurs questions aboutissent à une notation alternative. Au départ, cette exigence a souvent rencontré une vive réticence. Par la suite, il s'est avéré rapidement que cette formule conduit à une nette amélioration de la qualité du matériel d'évaluation. Le recours à la notation binaire requiert, plus que tout autre, un soin extrême dans l'élaboration des déclencheurs et des révélateurs proposés aux élèves. Rejetant toute demi-mesure, l'évaluateur doit fixer clairement son but, établir avec minutie les consignes de correction et définir sans ambiguïté les critères de réussite.

Dans certains cas, le matériel sur lequel a porté nos investigations reposait sur des épreuves évaluées à l'aide d'échelles numériques ou qualitatives à plusieurs niveaux. Une analyse approfondie du contenu des épreuves nous a conduit à la dichotomisation a posteriori des résultats. La subdivision en items qui en résulte peut d'ailleurs s'effectuer à la suite d'une succession de dichotomisations, en faisant varier les critères de réussite.

Bien entendu, une notation dichotomique ne peut déboucher sur la communication de résultats bruts aux élèves. Celle-ci serait d'ailleurs très mal accueillie. Le caractère tranché des observations, qui sert parfaitement notre objectif, est incompatible avec le but poursuivi lors de la communication de notes, à savoir le renforcement des conduites d'apprentissage.

4. CHOIX D'UN CRITERE DE CONCORDANCE

Quelles que soient les précautions prises lors de la construction des déclencheurs et des révélateurs, le professeur ou l'expérimentateur ne dépasse guère le niveau des hypothèses, tant en ce qui concerne les buts visés par l'évaluation et les caractéristiques intrinsèques de l'activité qu'à l'égard des conditions d'exécution du travail.

La qualité du matériel n'est pas toujours en cause. Différentes variables influencent les résultats enregistrés : aptitudes particulières des élèves, circonstances dans lesquelles l'épreuve se déroule, etc. Un même révélateur provoque souvent des réactions très différentes selon les classes et le contexte dans lequel on le propose, par exemple en fonction de la sanction attachée à l'évaluation.

Par conséquent, nous ne souhaitons déterminer les items susceptibles d'apporter des informations cohérentes qu'a posteriori, lorsque les données ont été enregistrées. Chaque item ne représente que lui-même. Bien sûr, les élèves, pris individuellement, y ont réussi ou échoué, ce qui apporte, sur le plan de l'évaluation des

individus des renseignements non négligeables. Mais pour l'évaluation de la tâche elle-même, et surtout en raison du caractère même des effets didactiques qui ont pour objet des groupes, on ne peut distinguer les items apportant des informations substantielles et décisives que lorsque la conjonction des comportements de plusieurs d'entre eux provoque un regroupement des performances des élèves. L'item n'est pris en considération que lorsqu'il voit son comportement confirmé par celui d'autres items.

Sans espérer atteindre le degré de perfection des sciences dites exactes, puisque très souvent des contraintes temporelles l'obligent à agir rapidement, l'éducateur se doit de manifester à l'égard de ses observations une grande prudence.

Nous appelons "famille d'items" un groupe d'items dont la concordance permet d'invoquer l'intervention d'un facteur commun. Cette concordance consacre la validité de l'ensemble des items par rapport au critère ainsi trouvé. Les causes probables de ce comportement seront alors recherchées, par exemple les effets didactiques qui l'ont éventuellement provoqué. Quant aux autres items, ils sont le résultat de trop d'influences ambiguës ou contradictoires pour que les informations qu'ils véhiculent soient prises en considération dans l'immédiat.

DANS L'ANALYSE DES TACHES SCOLAIRES, UN ITEM QUELCONQUE N'EST SUSCEPTIBLE D'APPORTER DES INFORMATIONS UTILES QUE S'IL EST POSSIBLE D'EN DECOUVRIR D'AUTRES AUXQUELS LES MEMES ELEVES REAGISSENT DE LA MEME MANIERE DANS DES CIRCONSTANCES ANALOGUES.

Notre position est opposée à celle des constructeurs de tests. Lorsque ceux-ci observent une grande corrélation entre deux items, ils doivent en éliminer un, qui risque de faire double emploi. Notre objectif est très différent, car comme le dit MATALON (1965 - p. 18) :

"... Si on essaie de construire un instrument de mesure, on pourra éliminer un item dont la corrélation avec un autre est trop forte : la présence de cet item n'ajoute rien au pouvoir discriminatif de l'instrument. On pourra ainsi négliger un item auquel tous les sujets réagissent de la même façon. En revanche, si on s'intéresse aux comportements eux-mêmes, ces résultats pourront se révéler aussi importants que d'autres."

Quant à la qualité du matériel d'évaluation, nous rejoignons GAGNÉ (1967), qui déclare avec beaucoup de bon sens que deux questions valent mieux qu'une, afin d'assurer la précision de la mesure. Après tout, il est facile de rédiger une mauvaise question, qui ne mesure pas ce qu'elle est censée mesurer...

L'interprétation d'une "famille d'items" se fonde sur la validation de chaque item par les autres. Les aspects techniques de mise en évidence d'une "famille" ne doivent pas faire oublier les buts poursuivis par l'évaluateur. L'observation d'une "famille" d'items concordants n'est jamais une garantie absolue d'obtenir une information cohérente. Seule une analyse approfondie du contenu et de la forme des items, ainsi que du contexte, conduit à l'explication des phénomènes et garantit la validité de leur interprétation.

CHAPITRE III SCHEMAS D'ANALYSE DE L'EVALUATION INFORMATIVE

Dans ce chapitre, nous présentons de manière détaillée les différents schémas d'analyse des effets collectifs sur lesquels nos recherches se sont fondées.

Dans un premier temps, l'adéquation entre les objectifs fixés par l'éducateur et les résultats effectivement enregistrés retient toute notre attention. Le schéma exposé s'attache à une explication simple, rationnelle, logique, par une répartition non équivoque des données observées, et leur comparaison aux objectifs éducatifs poursuivis. Cette première approche, limitée volontairement à une population réduite telle que nous la rencontrons habituellement dans nos classes, est ensuite élargie à des groupes d'individus plus nombreux.

Par la suite, la recherche de techniques d'analyse plus fines, mais aussi plus difficiles d'accès, nous conduit vers la mise en évidence de structures hiérarchiques. Le caractère orienté des interventions pédagogiques nous fait abandonner les modèles hiérarchiques totaux (GUTTMAN, CARDINET) au profit des hiérarchies centrées sur des critères bien contrôlés.

C'est par l'étude comparative des entités pédagogiques de base, c'est-à-dire les classes, que nous terminons la présentation de nos schémas. Les différentes classes, représentatives ou non d'une population parent, sont comparées entre elles. Des effets collectifs engendrés par des structures institutionnelles sont ainsi recherchées.

La présentation de chacun des schémas est accompagnée de commentaires critiques et d'algorithmes d'utilisation de la technique.

I LA CORRESPONDANCE ENTRE UN NIVEAU ATTEINT ET UNE AMBITION ANNONCEE

1°) Présentation du problème : la catégotisation des items

A condition d'être caractérisés par une "famille" d'observations concordantes, les items peuvent se classer en trois catégories :

a)	b)	c)	d)
+	O	+	
ITEMS D'ENQUETE SUPERIEUR	ITEMS D'ENQUETE INFERIEURE	ITEMS DE DIFFERENCIATION O	INTERACTION
Items à % élevé	Items à % faible	Items à % moyen	

1. L'observation confirme les ambitions (a)

L'objectif est atteint. Le succès de l'initiative considérée s'en trouve consacré. Le pourcentage de réussite à la "famille d'items" est élevé. (ITEMS D'ENQUETE SUPERIEURE)

2. L'observation infirme les ambitions (b)

L'objectif n'est pas atteint. Les résultats donnent lieu à un constat de carence. Les moyens mis en oeuvre n'ont pas porté leurs fruits. Le pourcentage de réussite à la "famille d'items" est faible. (ITEMS D'ENQUETE INFERIEURE)

3. L'observation n'est pas de nature à caractériser le résultat (ITEMS MOYENS)

Pour chacun des items, une partie seulement de la population a atteint l'objectif. Tous ces items sont difficiles à interpréter, sauf lorsqu'ils se regroupent en "familles", c'est-à-dire lorsqu'on

peut mettre en évidence des collections qui opposent systématiquement deux sous-groupes d'élèves : d'une part, ceux qui maîtrisent parfaitement l'objectif, d'autre part ceux qui y sont complètement inaccessibles. (FAMILLE DE DIFFERENCIATION). Dans ce cas, on conclut à un effet positif de la didactique au bénéfice d'un sous-groupe de sujets.

Si des items ne se regroupent pas en "familles", aucune conclusion collective ne peut être tirée. Tout au plus est-il possible de les considérer comme générateurs d'hypothèses qu'il conviendra de vérifier soigneusement par la suite (ITEMS D'INTERACTION).

La répartition des items s'effectue donc à partir de deux constatations fondamentales :

1. le niveau de la performance, qui s'exprime par le pourcentage de réussite à l'item;
2. la concordance du comportement des items.

Pour les items à pourcentage de réussite extrême, la concordance est nécessairement toujours élevée. Seule, l'interprétation du contenu des items consacre la validité des observations.

Pour les items à pourcentage de réussite moyen, la différenciation implique la présence d'un groupe d'items suffisamment concordants. La différenciation constitue donc par elle-même un facteur déterminant pour la mise en évidence d'informations fiables.

De nombreux auteurs se sont intéressés aux résultats enregistrés aux différents items d'une épreuve, mais leur objectif était tout différent du nôtre, et même le plus souvent radicalement opposé.

a) Exploitation du pourcentage de réussite

Le pourcentage de réussite aux items a souvent été galvaudé par les auteurs. Ceux-ci ne l'utilisent qu'en vue de la construction d'épreuves normées ou de la constitution de banques d'items. La quête d'indices absolus et définitifs les conduit à des prises de position qui les éloignent fort de nos préoccupations.

Pour DE LANDSHEERE (1972), l'indice de facilité des questions est utile dans le cadre de la réalisation d'un test ou d'une épreuve équilibrée, mais il faut disposer d'un grand nombre de sujets (au moins 100...).

RIGAUX (1962) subdivise les items en trois grandes catégories:

1. Les items à indice de difficulté "élevé", c'est-à-dire ceux qui ont été réussis par un nombre infime de sujets;
2. Les items à indice de difficulté "moyen", c'est-à-dire ceux qui ont été réussis par à peu près la moitié de l'échantillon;
3. Les items à indice de difficulté "faible", c'est-à-dire ceux qui ont été réussis par la plupart des sujets.

Il considère les items de la seconde catégorie comme les meilleurs, car son objectif exclusif reste la discrimination entre sujets forts et sujets faibles.

PIDGEON et YATES (1968) conseillent, lors de l'analyse des items en vue de la construction d'un test, de rejeter toute question qui se situe en dehors de l'intervalle 30 % - 70 % !

La position de THOMAS (1954) est assez exceptionnelle pour que nous la soulignions. Il suggère en effet de retenir les items à pourcentage de réussite extrême pour juger de la qualité des moyens d'enseignement et du matériel utilisé en classe (p. 12) :

" Le professeur peut se concentrer sur lui-même et sur ses propres procédés. C'est-à-dire qu'il lui est possible, au travers des données d'évaluation, de juger l'efficacité des activités d'enseignement et du matériel utilisé en classe. Par exemple, si à l'aide d'un test ou d'une échelle de classement, l'enseignant découvre que toute la classe a lamentablement échoué, alors la méthode et le matériel utilisés par le professeur se sont apparemment révélés inadéquats pour tous les étudiants. Si, d'autre part, les tests font apparaître que la plupart des étudiants ont atteint les objectifs, nous concluons que les méthodes d'enseignement convenaient vraisemblablement à la majorité. Mais pour les quelques étudiants qui ont échoué, les méthodes et le matériel étaient inadéquats."

Malgré son souci de retenir les informations relatives aux effets de l'action éducative, THOMAS ne peut s'empêcher in fine de s'abandonner à la tentation d'une évaluation de l'individu...

b) L'opposition entre deux sous-groupes d'élèves

La discrimination des élèves a beaucoup préoccupé les constructeurs de tests dans le cadre de l'analyse des items. Ils cherchent à opposer deux groupes extrêmes, la plupart du temps en négligeant les élèves qui obtiennent une note moyenne, afin de déclarer certains items "discriminatifs". Mais la discrimination est souvent recherchée pour elle-même : les questions discriminatives sont automatiquement les "bonnes questions", quelles que soient leurs caractéristiques. Cette tendance suspecte, qui amène les utilisateurs à conférer des qualités à un matériel parce qu'il présente des indices statistiques élevés, doit être résolument combattue. Seule une interprétation minutieuse permet de confirmer la validité des informations obtenues.

DE LANDSHEERE (1972) parle ainsi d'un indice d'efficacité, qui révèle dans quelle mesure une question discrimine les élèves forts des élèves faibles.

PIDGEON et YATES (1968) proposent une procédure destinée à mesurer l'efficacité (efficiency) des items, mais ils ne suggèrent aucun critère qui préciserait la signification de leur indice.

a) On divise la classe en 3 groupes : supérieur, moyen, inférieur, sur base des scores à l'ensemble du test.
b) On calcule le pourcentage de réussite de chaque item dans chacun des trois groupes.
c) On obtient un indice satisfaisant en soustrayant la proportion de réussite chez les inférieurs de la proportion de réussite chez les supérieurs

DAVIS (1949), à côté d'un indice de difficulté calculé à partir du pourcentage de réussite, construit un indice de discrimination, qui indique avec quelle certitude un item donné discri-

mine le groupe fort du groupe faible. Se fondant sur les études de KELLEY (1939), il compare les pourcentages de réussite des 27 % supérieurs de la population et des 27 % inférieurs. Ces pourcentages sont introduits dans une table qui fournit l'indice de discrimination correspondant, lequel, d'après DAVIS, n'aurait qu'une valeur indicative.

KELLEY (1939), dont il vient d'être question, assortit la démonstration qui fournit le critère de 27 % de deux conclusions intéressantes :

1. Les items les plus fiables sont en connexion avec des groupes obtenant 50 % de bonnes réponses.
2. Des modifications à partir de sa procédure - qui est bien entendu recommandée - sont théoriquement souhaitables, surtout si l'on se trouve devant des distributions non normales ou des items de difficulté non adaptée aux groupes observés. Il ajoute cependant : " Though we may believe that they are hardly feasible practically...".

En réponse aux propositions de KELLEY, RIGAUX (1961) a constaté que les indices de difficulté obtenus à partir des 27 % supérieurs et inférieurs diffèrent peu des indices établis sur l'échantillon total. Il conteste la méthode préconisée par DAVIS dans la mesure où les distributions observées ne sont pas normales. RIGAUX recourt quant à lui à la Théorie de l'Information pour déterminer la valeur discriminative des items.

Notre conception de la différenciation, qui discrimine effectivement deux sous-groupes parmi les élèves, s'éloigne fortement de celles que nous venons de décrire. Les élèves eux-mêmes ne sont pas mis en cause, mais bien la didactique qui a abouti à leur discrimination. Si l'objectif avéré de l'enseignement est la sélection des individus les plus aptes, une différenciation consacre la validité de la didactique. Inversément, une éducation qui se déclare non sélective mettrait tout en oeuvre pour contrecarrer les causes d'une différenciation.

2°) Moyens techniques proposés

A. Effectifs compris entre 10 et 30 élèves

1. Présentation de la technique

Pour définir des familles d'items en termes d'objectifs d'enseignement, le critère le plus simple est constitué par la difficulté observée, c'est-à-dire le pourcentage de réussite obtenu à chaque item.

En termes de probabilités, nous admettrons que la réussite ou l'échec, caractérisés l'une par un pourcentage de réussite élevé, l'autre par un faible pourcentage de réussite, ne proviennent pas du hasard, mais bien d'un effet, positif ou négatif, de l'enseignement dispensé.

A l'inverse, les autres items, c'est-à-dire ceux qui donnent des résultats voisins de 50 % de réussite, sont a priori assimilables à des événements de nature aléatoire.

La loi binomiale des probabilités nous permet de construire une table d'intervalle autour de 50 % de réussite en fonction du nombre d'élèves observés. Nous avons construit notre table pour des populations variant de 10 à 30 élèves, qui correspondent à la taille habituelle des classes de nos régions.

TABLE D'INTERVALLE AUTOUR DE 50% DE REUSSITE

N	intervalle	N	intervalle
10	2 - 8	21	6 - 15
11	3 - 8	22	7 - 15
12	3 - 9	23	7 - 16
13	3 - 10	24	7 - 17
14	4 - 10	25	8 - 17
15	4 - 11	26	8 - 18
16	4 - 12	27	9 - 18
17	5 - 12	28	9 - 19
18	5 - 13	29	9 - 20
19	6 - 13	30	10 - 20
20	6 - 14	31	10 - 21

Cette table donne, pour des effectifs N, le minimum et le maximum de réponses exactes qu'un item peut admettre pour se situer dans l'intervalle (limites incluses).

Notre schéma d'interprétation n'admet pas l'existence d'un effet collectif en deçà ou au delà de la limite de l'intervalle de confiance. Les écarts par rapport au minimum ou au maximum sont aléatoires et relèvent d'un ensemble de variables parasites. Nous n'entendons pas sous-estimer l'intérêt de ces influences, mais en évaluation informative, il ne peut être question d'en rendre responsable une didactique intentionnelle.

Cependant, ces conclusions quant au succès ou à l'échec de l'action éducative ne sont valides qu'à la condition de disposer de plusieurs items relatifs à la même compétence et assortis de la même réussite.

En vertu de la loi binomiale des probabilités, les items situés dans l'intervalle autour de 50 % de réussite sont a priori suspects, puisque le risque est grand que des facteurs indépendants de la didactique entraînent la réussite ou l'échec. Nous n'accordons de crédit qu'à une collection d'items présentant des comportements communs, donc des distributions telles que les élèves se différencient en deux groupes opposés, comme dans le test de classement de CARDINET (1972).

Le problème revient à identifier, parmi les items dont la proportion de réussite est comprise dans l'intervalle de confiance autour de 50 %, ceux qui, par la similitude des distributions de leurs résultats, permettent une interprétation.

Nous avons proposé par ailleurs (1) un algorithme pour la détection des items de différenciation :

1° Calcul de la note "provisoire" (NP) par élève sur les seuls items situés à l'intérieur de l'intervalle de confiance autour de 50 % de réussite.

2° Calcul de la note de l'élève moyen ($\bar{x}e$) sur les notes provisoires.

$$\bar{x}e = \frac{\Sigma(NP)}{N}$$

---

(1) VANDEVELDE et LEBRUN Essai d'évaluation scolaire informative Revue belge de psychologie et de pédagogie n° 152 Déc. 1975 pp 77 - 98

La note de l'élève "moyen" permet de séparer les élèves en deux groupes : les élèves "supérieurs" à la note moyenne et les élèves "inférieurs" à celle-ci. Cette distribution des élèves en deux catégories n'est pas définitive. Il arrive que des notes proches de la moyenne soient placées à tort dans l'un des deux groupes, mais la suite de la technique rectifie les éventuelles erreurs de répartition. Les termes "supérieurs" et "inférieurs" doivent être pris dans leur sens strict de situation par rapport à l'élève moyen.

3° Calcul du pouvoir de différenciation de chaque item.

Il s'agit de distinguer, parmi l'ensemble des items dont le pourcentage de réussite est assimilable à 50 %, ceux qui différencient les élèves en se réunissant en une "famille". L'existence du pouvoir de différenciation éventuel d'un item se manifeste par la concordance entre cet item et la distribution des élèves en deux groupes. Le degré de concordance est calculé par l'INDICE DE DIFFERENCIATION SPECIFIQUE (IDS).

$$IDS = \frac{\Sigma+s}{nes} - \frac{\Sigma+i}{nei}$$

où $\Sigma+s$: nombre de + obtenus par les élèves "supérieurs",
nes : nombre d'élèves "supérieurs",
$\Sigma+i$: nombre de + obtenus par les élèves "inférieurs",
nei : nombre d'élèves "inférieurs".

Tous les items dont l'IDS atteint ou dépasse le seuil de .50 sont admis dans la "famille" des items de différenciation.

4° Calcul de la note de différenciation (ND) pour chaque élève, sur l'ensemble des items de différenciation.

L'interprétation d'une "famille" d'items de différenciation se fonde sur l'opposition entre un groupe d'élèves pour lequel le pourcentage de réussite est assimilable au maximum, et un autre groupe pour lequel la didactique s'est avérée inefficace.

Contrairement à la pratique admise en évaluation classique, nous refusons d'additionner les notes appartenant à des distributions incohérentes. Les résultats ne sont pris en considération qu'à partir du moment où l'examimateur a obtenu des items dont les réponses se distribuent de manière analogue. Dans cette procédure, c'est l'observation de phénomènes collectifs qui prime, et non plus l'évaluation de compétences individuelles.

Il est possible de dépouiller très rapidement une matrice d'évaluation informative grâce à des procédés de marquage (annexe 3). Le recours à l'ordinateur permet cependant d'alléger sensiblement le travail.

2. Commentaires sur la technique proposée

a) Recours à la loi binomiale des probabilités pour construire la table d'intervalle

Dans le cas d'une situation dichotomique, c'est-à-dire lorsque chaque individu possède l'un ou l'autre de deux caractères opposés, on applique la loi binomiale - parfois appelée "loi d'alternative répétée" (1) - lors d'un prélèvement au hasard dans une population infinie d'individus.

Si le caractère aléatoire d'une réponse alternative se vérifiait, la probabilité de répondre correctement serait de .50. C'est donc à 50 % de réussite que se situe la plus grande probabilité pour obtenir des réponses provoquées par des influences incontrôlées. Plus le pourcentage de réussite à un item donné s'éloigne de 50 %, plus diminue la probabilité liée au hasard qui lui est associée. Lorsqu'il s'agit de jeux de hasard, le jet de pièces de monnaie par exemple, on peut soupçonner le truquage des pièces lorsque cette probabilité devient très faible, en fait lorsqu'elle tombe sous un seuil critique donné.

(1) DAGNELIE Théorie et méthodes statistiques 1970

L'intervalle de confiance autour de 50 % de réussite, à l'intérieur duquel on considère que l'effet du hasard a pu intervenir, est obtenu en additionnant les probabilités afférentes aux diverses situations observables autour de 50 %, jusqu'à concurrence d'un seuil déterminé.

Les items qui satisfont à notre hypothèse nulle - H_o = 50 % - ne présentent aucun intérêt par eux-mêmes, puisque livrés a priori à des influences incontrôlées. Nous nous interrogeons par contre sur les raisons qui provoquent le rejet de certains items. Nous rejoignons ici l'usage de l'hypothèse nulle que défend BENZECRI (1973 - Vol. I, p. 22) :

" En termes de statistique, nous pourrons dire que nous avons choisi l'indépendance comme hypothèse nulle H_o. Mais l'usage que nous faisons de l'hypothèse nulle est tout différent de celui que l'on en fait habituellement en statistique appliquée : il ne s'agit pas de savoir s'il faut, ou non, rejeter l'hypothèse nulle (s'il faut la retenir, c'est que les données sont dénuées de signification, et donc sans intérêt!). Mais s'il faut la rejeter, c'est seulement le signe que l'on peut faire quelque chose, ce n'est nullement un résultat en soi; nous voulons voir apparaître aussi clairement que possible comment nos données diffèrent de H_o, dans quelle direction elles s'écartent de l'indépendance."

En pédagogie, on considère qu'un effet lié à la didactique est intervenu lorsque le seuil des 95 % est atteint, c'est-à-dire 47,5 % de part et d'autre des 50 %. Il existe des tables qui donnent les valeurs correspondant à ce seuil. Nous avons employé la table communément appelée "Table de Hald" pour construire la nôtre. Le nombre des élèves testés a une influence sur l'intervalle de confiance déterminé. Nous discuterons ce point plus loin.

b) <u>Sommation des notes enregistrées à 50 % de réussite</u>

Nous nous sommes refusé à admettre la sommation des résultats par élève sur l'ensemble des items constituant l'épreuve. Toutes les techniques destinées au classement des élèves font cette somme. Nous appelons d'ailleurs la note qui en résulte la "NOTE ORDINALE" (NO).

Cependant, pour des raisons pratiques, nous effectuons une sommation des résultats, mais uniquement sur les items à 50 % de réussite. Nous appelons la note obtenue la "NOTE PROVISOIRE" (NP). Par la suite, nous établissons deux groupes : les items de différenciation - apportant une information sur des sous-groupes dans la classe - et les items d'interaction. Par définition, les items d'interaction ont chacun un comportement différent, et nous pouvons supposer qu'en probabilité leurs effets vont s'annuler. D'autre part, une différenciation éventuelle dans l'épreuve doit normalement se trouver en germe dans la note provisoire.

c) Dichotomisation des notes provisoires

Le choix d'un critère s'impose pour dichotomiser la distribution des résultats et la rendre compatible avec la suite des opérations. La moyenne arithmétique s'est avérée l'instrument le plus favorable. Cette moyenne n'est que rarement un nombre entier. Elle permet de distinguer facilement les élèves supérieurs des élèves inférieurs. Si la moyenne est un nombre entier, il suffit de placer ceux qui l'obtiennent parmi le groupe supérieur. Les calculs ultérieurs permettront d'entériner cette hypothèse.

Les élèves "moyens", pour lesquels il est très difficile de discerner l'impact réel des effets de la didactique, sont particulièrement sensibles à la technique. Ou ils migrent nettement vers l'un ou l'autre groupe d'élèves, ce qui permet de les cataloguer et de leur appliquer les traitements appropriés, ou ils conservent une position moyenne et restent réfractaires aux moyens d'investigation utilisés. D'autres instruments, plus fins, devront être envisagés, qui détermineront avec précision leurs forces et leurs faiblesses en vue de mettre en oeuvre les correctifs adéquats.

La note provisoire et la moyenne arithmétique ne constituent rien d'autre que des artifices de calcul. L'approche par l'évaluation hiérarchique que nous verrons plus tard, en se basant sur toutes les interconcordances entre chaque paire d'items, même en dehors de l'intervalle de confiance autour de 50 % de réussite, évite le recours à une dichotomisation provisoire. Cette technique n'est malheureusement pas applicable à de petits nombres d'élèves. De plus, elle nécessite de longs détours et le recours à l'ordinateur.

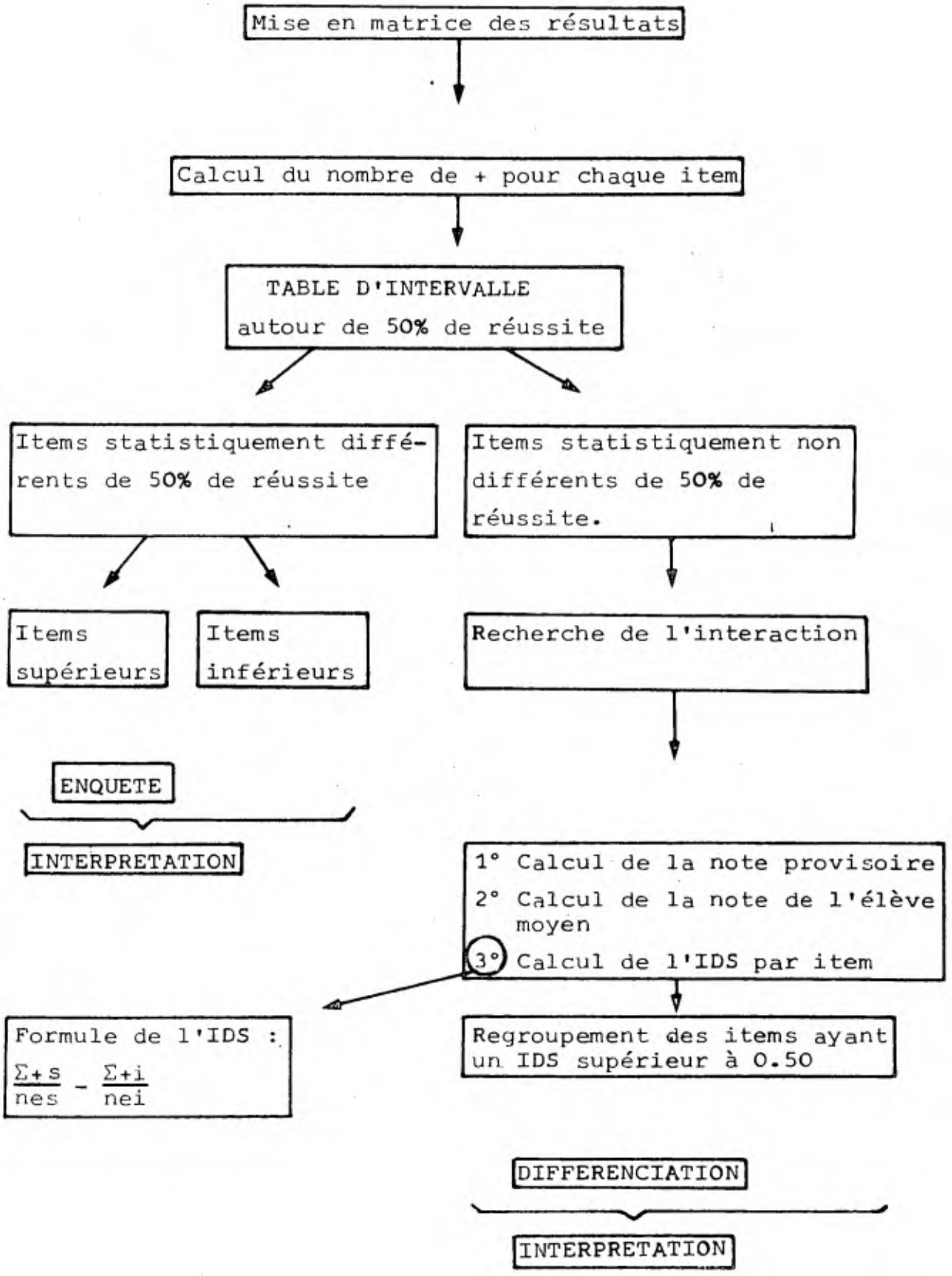
d) L'indice de différenciation spécifique (IDS)

Le calcul de l'IDS, extrêmement simple, ne peut rebuter les plus allergiques aux manipulations de chiffres. Sa logique interne est irréfutable et correspond parfaitement à la représentation typique de la corrélation. Il s'agit, ni plus ni moins, de soustraire la proportion de réussite des élèves dits "inférieurs" (en opposition avec le critère choisi) de la proportion de réussite des élèves dits "supérieurs".

L'IDS varie de +1 à -1, comme les indices de corrélation, et nous verrons plus loin que sa familiarité avec l'indice de corrélation de point (φ) est très grande.

C'est de manière tout empirique que le critère de .50 associé à l'IDS a été fixé pour l'acceptation d'un item dans la différenciation. L'analyse de l'évolution de la binomiale nous permettra de mettre en rapport avec ce critère une interprétation dans laquelle la proportion de .46 prévaudrait. Nous continuons cependant de proposer .50 aux professeurs. Cette proportion est psychologiquement mieux adaptée. De toute manière, elle est statistiquement plus sévère.

3. Tableau schématique de la technique.



B. Extension à des populations plus vastes

Les moyens préconisés jusqu'ici se justifient dans le cadre d'une utilisation limitée à la classe et sans autre ambition que de contrôler l'évolution de la didactique quotidienne.

Pour l'extension à des effectifs plus élevés et en cas d'objectifs plus larges, une adéquation des techniques doit être envisagée.

1. Présentation de la technique

a) L'intervalle de confiance autour de 50 % de réussite

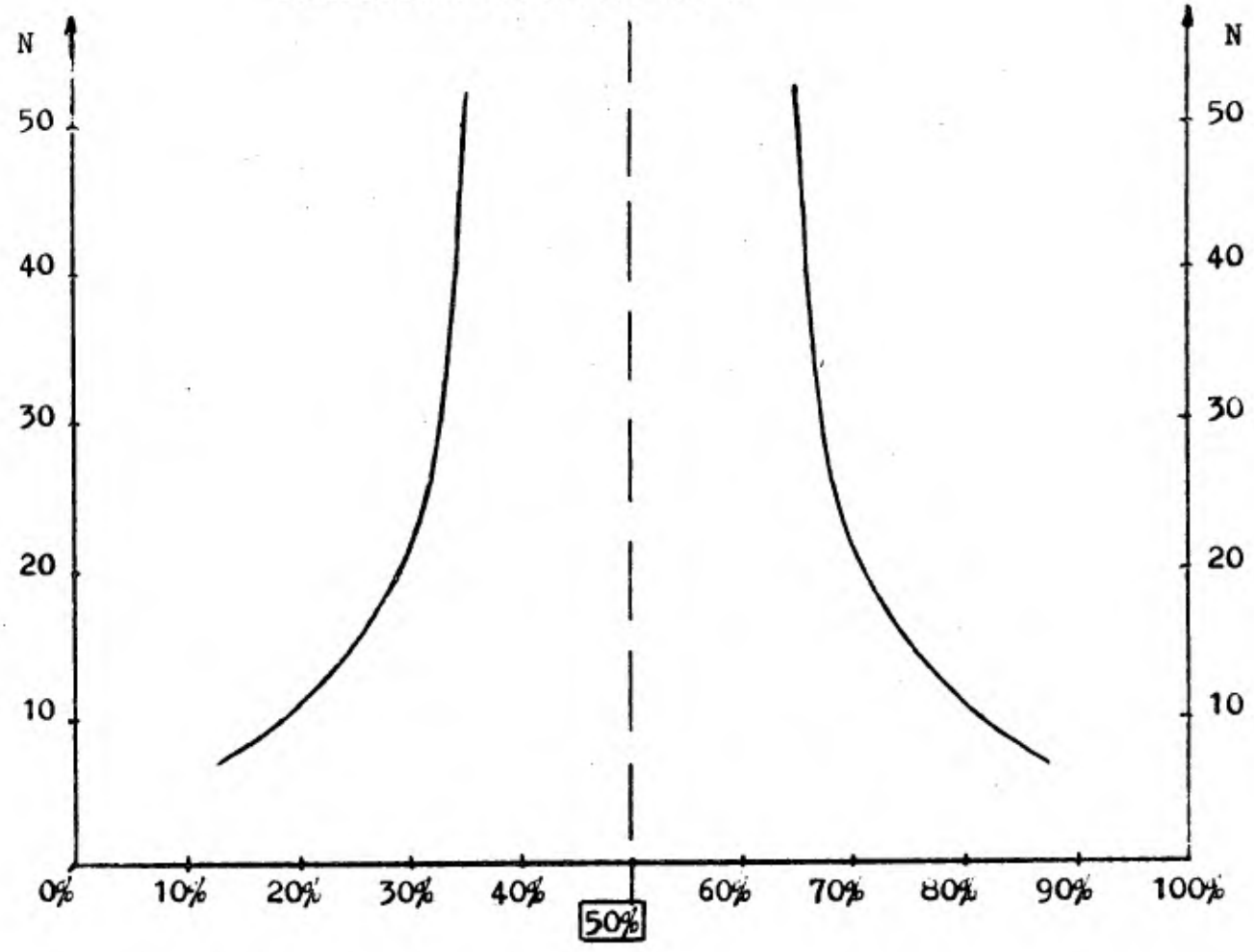
La répartition des items en catégories repose en fait sur les principes suivants :

- tout item supérieur à l'intervalle autour de 50 % de réussite est assimilable à une réussite maximale ;
- tout item inférieur à l'intervalle autour de 50 % de réussite est assimilable à une réussite nulle.

Le choix que nous avons fait de la binomiale pour mettre en évidence ces types d'items, largement justifié dans le cadre restreint d'une classe, s'avère peu indiqué lorsque le nombre d'élèves devient très élevé.

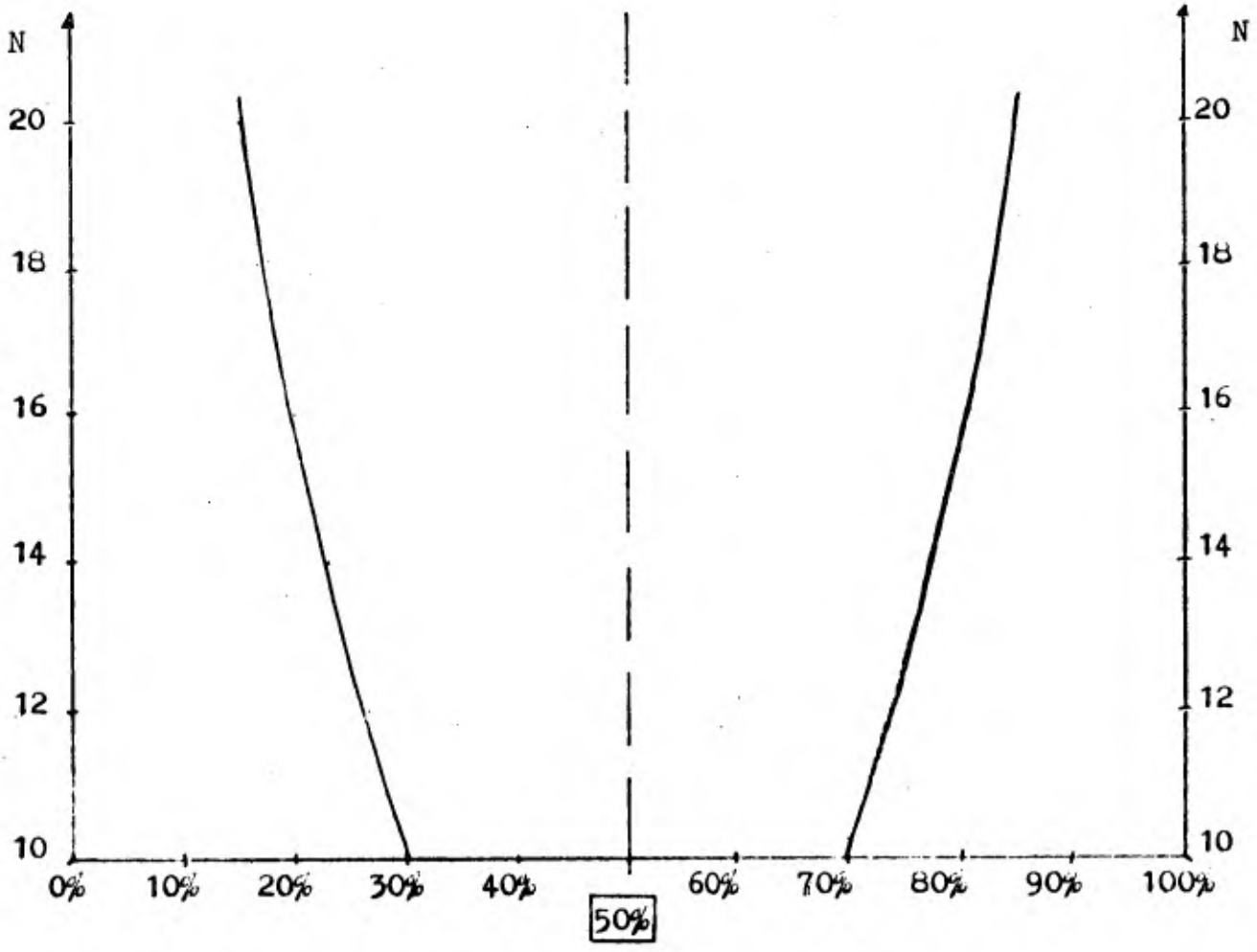
Une étude de la variation, dans le tableau ci-après, nous montre que si l'intervalle croît en valeur absolue avec le nombre d'élèves, le pourcentage s'amenuise au point de remettre en question les principes énoncés plus haut. Il faut donc déterminer un intervalle indépendant du nombre d'élèves, et l'on peut situer les limites de cet intervalle aux environs de 25 % et de 75 %.

EVOLUTION DE LA BINOMIALE

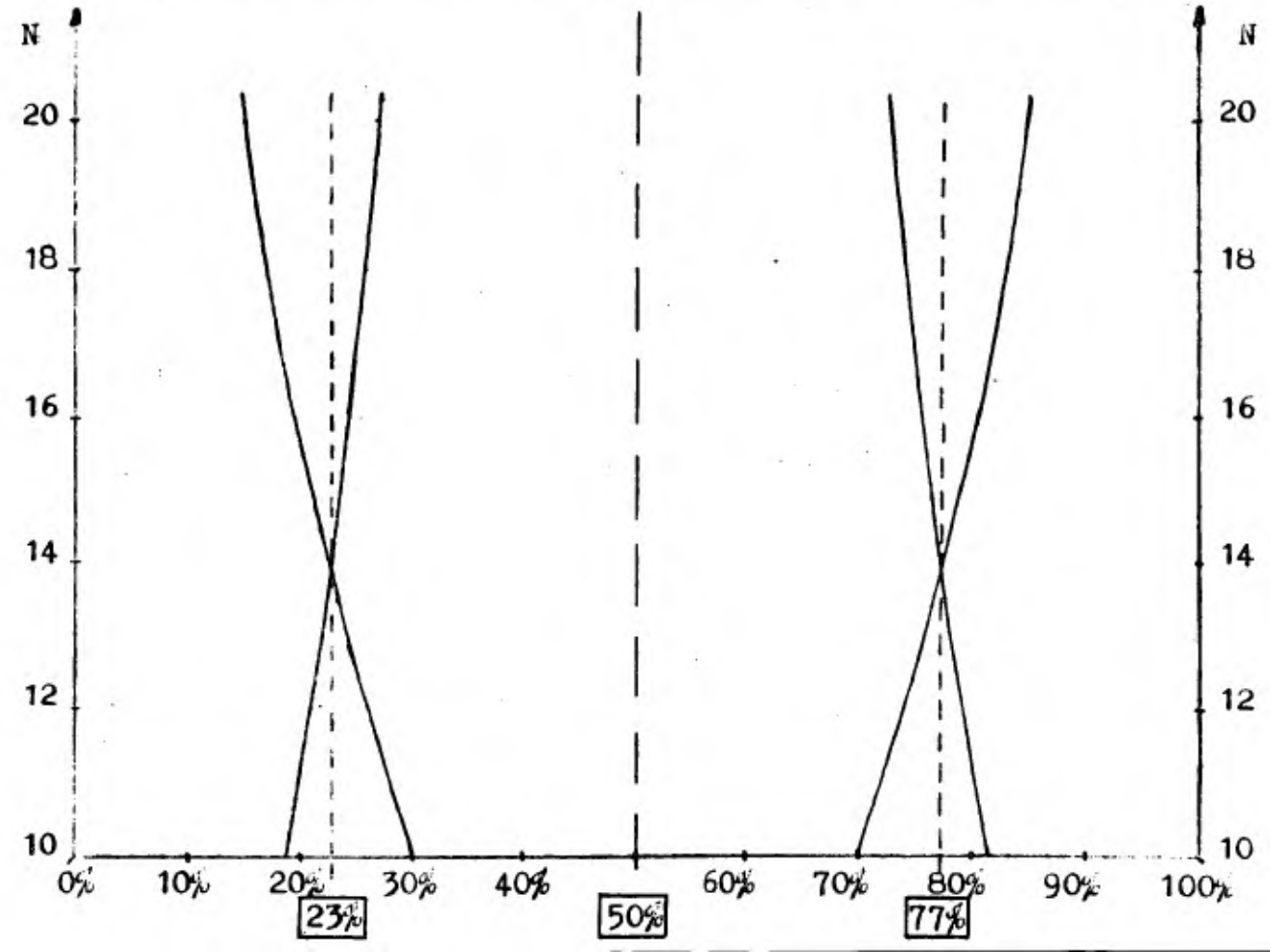


Le développement qui suit ne peut en aucun cas être considéré comme une démonstration statistique débouchant sur la fixation d'un intervalle péremptoire. Il constitue cependant une observation utile, et nous n'avons pas craint de nous rallier à ses conclusions, lesquelles correspondent davantage à des constatations empiriques qu'à des déductions purement spéculatives.

S'il est possible de déterminer un intervalle de confiance autour de la proportion de 50 % de réussite à partir de la loi binomiale, on peut appliquer le même raisonnement pour les proportions assimilables à 100 % ou à 0 % de réussite. Nous avons indiqué les résultats obtenus dans le tableau suivant :



Confondons les deux tableaux précédents dans le tableau que voici :



a) pour $N < 14$, il y a recouvrement des intervalles, c'est-à-dire qu'un item peut être simultanément assimilable à 50 % et à l'un des extrêmes ! Dans ce cas, la technique basée sur la binomiale surestime l'intervalle autour de 50 %. Elle s'avère trop sévère quant il s'agit de conclure que l'on a bien procédé à une enquête.

b) pour $N > 14$, il existe une frange d'items qui ne sont assimilables ni à 50 %, ni au maximum ou à zéro. La technique conclut abusivement à l'enquête lorsque le nombre d'élèves croît

c) pour $N \approx 14$, nous rejoignons les conclusions formulées dans les principes énoncés précédemment. Un item est alors parfaitement caractérisé par son pourcentage de réussite.

Nous allons extrapoler cette dernière observation pour en tirer les conclusions suivantes :

- tout item au pourcentage de réussite supérieur à 77 % est assimilable à 100 % ;
- tout item au pourcentage de réussite inférieur à 23 % est assimilable à 0 % ;
- tout item au pourcentage de réussite compris entre 23 % et 77 % est assimilable à 50 %.

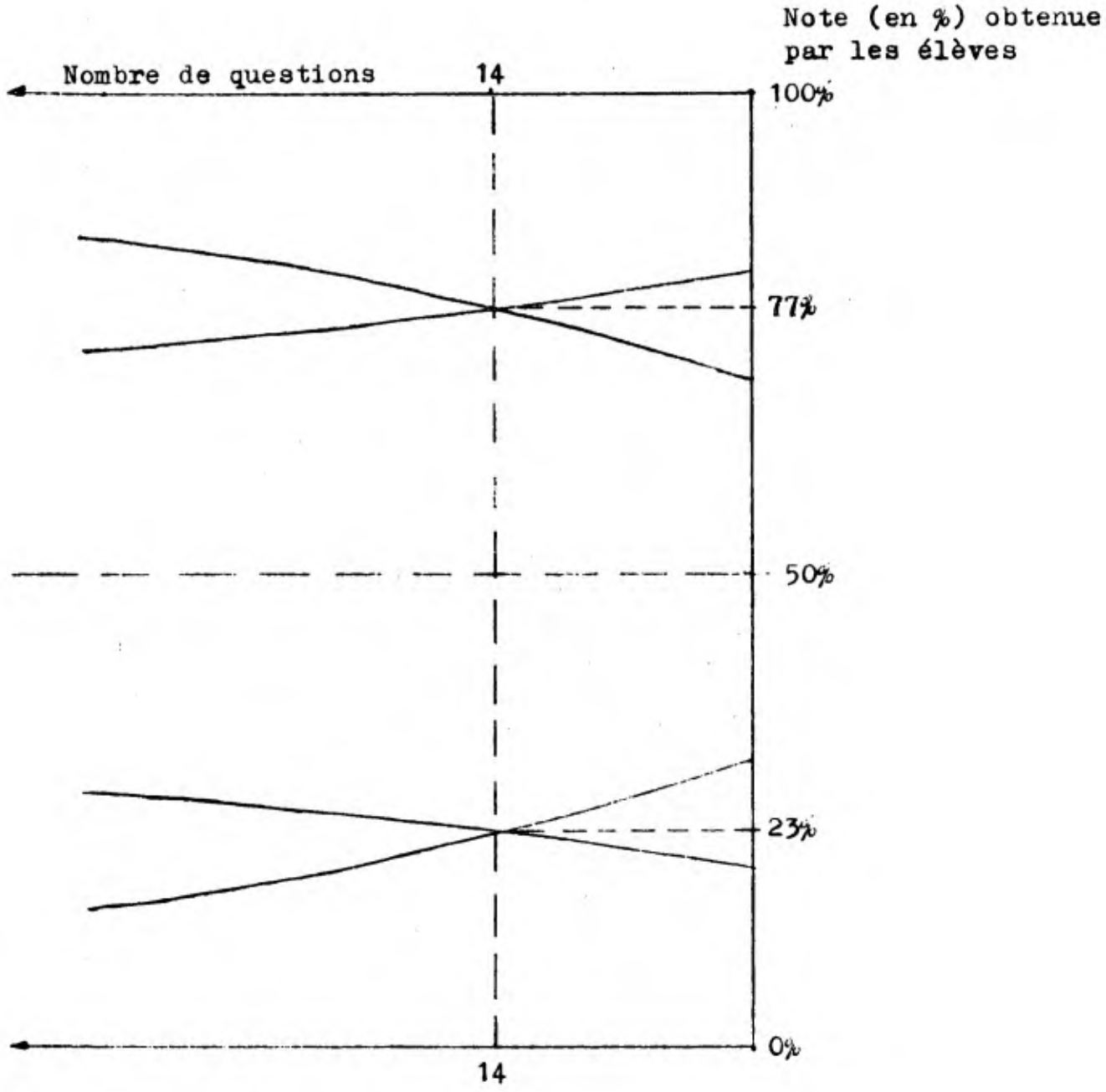
Nous pouvons ainsi catégoriser les items quel que soit le nombre d'élèves. En comparant plusieurs centaines de matrices analysées selon les deux critères (intervalle variable en fonction du nombre d'élèves, et intervalle fixé à 23 % - 77 %), nous avons constaté que les différences étaient minimes et négligeables pour des effectifs compris entre 10 et 30 élèves par classe.

Nous continuons à utiliser la technique à intervalle mobile dans l'étude des classes à effectif normal. L'application de l'intervalle de confiance fixe permet d'analyser des résultats

portant sur un nombre indéterminé d'élèves, c'est-à-dire pour des échantillons de classes, ou dans des classes comprenant un nombre d'élèves supérieur à 30, comme il en existe aujourd'hui dans de nombreux pays.

b) Le critère d'acceptation des items de différenciation

Nous avions fixé empiriquement le seuil d'acceptation d'un item de différenciation à .50. Nous pouvons appliquer le raisonnement que nous avons tenu pour la catégorisation des items en extrapolant les observations concernant la loi binomiale, mais en considérant cette fois les notes obtenues par les élèves.



Pour une épreuve de différenciation de 14 questions, les notes des élèves se répartissent en trois catégories :

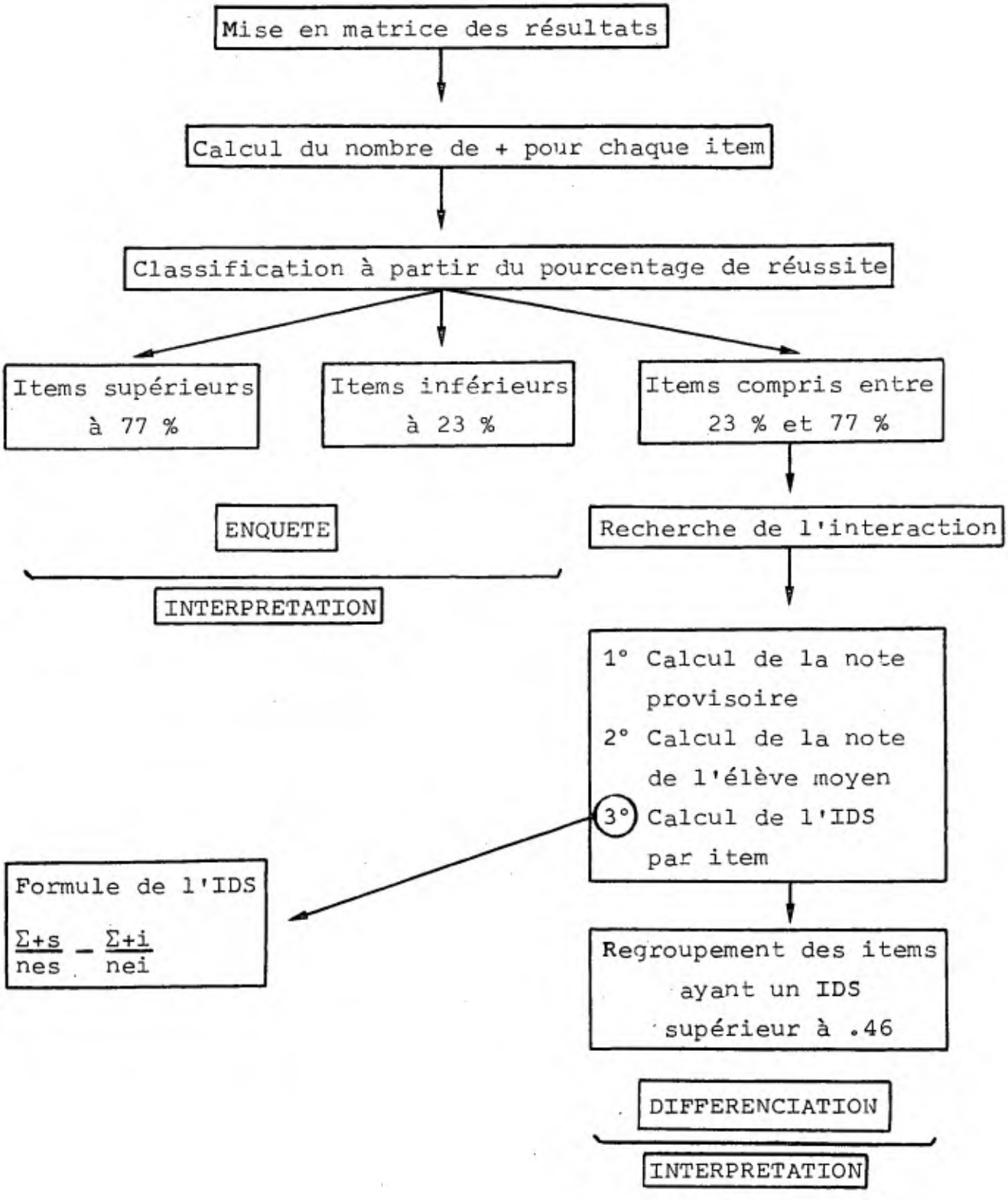
- les notes assimilables au maximum de réussite (> 77 %) ;
- les notes assimilables à zéro (< 23 %) ;
- les notes assimilables à 50 % de réussite (comprises entre 23 % et 77 %).

Sur l'ensemble des notes observables, nous retenons pour la différenciation 23 % de notes supérieures et 23 % de notes inférieures, c'est-à-dire au total 46 % de notes. Le comportement linéaire de l'indice de différenciation spécifique (IDS) nous autorise à retenir ce seuil de 46 % pour considérer qu'un item permet la différenciation des élèves.

Le seuil de 46 % est applicable à l'étude approfondie de populations nombreuses dans le but de mettre en évidence des effets scientifiquement observables. Il est forcément tributaire de la technique utilisée pour calculer la concordance. En effet, si le comportement linéaire de l'indice de différenciation spécifique nous empêche de recourir aux tests paramétriques, il nous permet cependant de confronter directement les résultats enregistrés aux représentations théoriques, sans effectuer de calculs complexes ni consulter de tables.

Il ne nous a pas semblé utile de modifier le critère de .50 qui avait été proposé pour l'analyse d'une classe isolée.

2. Tableau schématique de la technique



C. Conclusion

La correspondance entre un niveau atteint et une ambition annoncée constitue indiscutablement la pierre angulaire de l'évaluation informative. Les instruments que nous avons proposés sont très simples, peut-être même grossiers... C'est pourtant ce qui les rend accessibles à la masse des éducateurs et leur permet d'en tirer un profit appréciable. Nous trouvons ce schéma élémentaire à la base de toutes les observations que nous développons dans les annexes. Nous avons constaté une nette modification dans l'attitude des enseignants lorsqu'ils ont admis les principes de ce schéma : ils manifestent notamment une grande prudence dans le maniement et l'exploitation de leurs outils d'évaluation. Le plus difficile consiste toujours à leur faire comprendre que l'évaluation de l'élève ne peut suffire en éducation.

L'évaluation hiérarchique, dont il sera bientôt question, aura quant à elle des prétentions scientifiques plus affirmées.

II LA MISE EN EVIDENCE DE HIERARCHIES

1°) Présentation du problème : le classement des activités

Un des principaux avantages de la recherche d'une correspondance entre un niveau atteint et une ambition annoncée réside dans l'absence d'hypothèses préalables aux opérations d'analyse des résultats.

Lorsque nous nous tournons vers les structures hiérarchiques, notre but est de rechercher des informations utiles dans le cadre de l'action éducative. Les objectifs que nous poursuivons sont semblables à ceux que MATALON (1965) assigne à l'analyse hiérarchique (p. 16) :

" L'analyse hiérarchique met en évidence une structure très générale sous-jacente dans tous les cas où on cherche à construire une variable ordinale à partir de données binaires."

Cependant, lors de l'analyse des résultats de l'action éducative, il ne peut être question de laisser se constituer des hiérarchies dont les fondements ne seraient pas parfaitement contrôlés. Aussi avons-nous rompu d'emblée avec les techniques de mise en évidence de structures hiérarchiques totales, telles qu'elles sont conçues dans les échelles d'attitudes (GUTTMAN) et dans les sciences de la nature, où l'observation reste l'objectif ultime, pour nous tourner vers une hiérarchisation centrée sur un CRITERE.

Dans l'absolu, le critère est "ce qui sert de base à un jugement" (ROBERT). En évaluation informative, un CRITERE HIERARCHIQUE est défini par un ensemble de notions dont la cohésion est particulièrement forte. Ce critère s'associe à d'autres notions, moins consistantes, mais qui présentent une concordance suffisante pour former une hiérarchie.

Le critère sera donc le point de référence permanent, d'abord pour la constitution technique de la hiérarchie, ensuite pour son interprétation. L'évaluation informative, en s'attachant aux effets collectifs de l'éducation, impose un certain nombre de

contraintes techniques et formelles pour la détermination des critères susceptibles de servir de base à des hiérarchies.

Nous avons vu précédemment qu'une information cohérente, fiable, ne pouvait se fonder que sur une "famille d'items". Cette "famille" se compose nécessairement de plusieurs items concordants, mais nous avons émis des réserves sur la signification de cette concordance lorsque le pourcentage de réussite aux items s'éloigne de 50 %. Nous avons aussi insisté sur le fait qu'une "famille d'items" n'est exploitable que si des caractéristiques objectives (contenu, forme, contexte,...) ressortent de l'interprétation.

Nous avons étudié de très nombreux critères dans des cadres variés, et nous avons effectivement constaté que seuls les critères qui présentaient ces particularités servaient de base à des hiérarchies pédagogiquement acceptables (annexes 1, 4 et 5).

Trois conditions caractérisent donc un "CRITERE HIERARCHIQUE" :

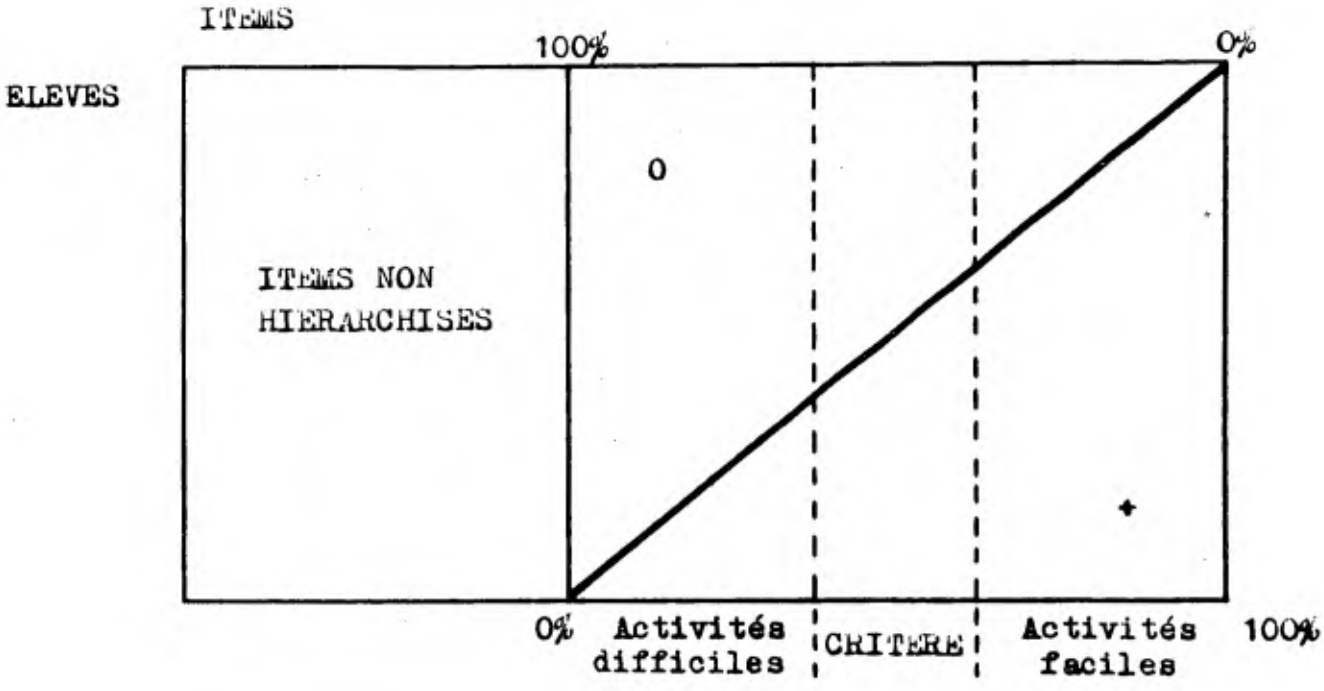
1°) il est constitué d'une "famille" d'items concordants ;

2°) ces items ont un pourcentage de réussite comparable et relativement peu éloigné de 50 % ;

3°) les propriétés de ces items sont interprétables, on peut les définir globalement.

Pour répondre à notre souci de n'utiliser que des techniques compréhensibles aussi bien par l'éducateur peu habile à manier des données statistiques que pour le chercheur spécialisé, nous nous sommes orienté vers un schéma plus logique que probabiliste, afin de mettre en évidence une hiérarchie autour d'une "famille-critère".

Le critère ayant été choisi, on recherche dans l'épreuve tous les items qui se hiérarchisent autour de lui.



La matrice constituée par tous les items hiérarchisés se décompose comme suit :

a) les items plus faciles que le "critère" (% de réussite supérieur à celui de l'item le plus facile du critère). Il s'agit le plus souvent de prérequis ou d'activités préparatoires.

b) les items plus difficiles que le "critère" (% de réussite inférieur à celui de l'item le plus difficile du critère). L'accessibilité limitée de ces items les réserve à une minorité d'individus, lesquels doivent préalablement maîtriser le critère et a fortiori les prérequis.

c) les items à pourcentage de réussite comparables à celui du "critère" (% de réussite compris entre celui de l'item le plus facile et celui de l'item le plus difficile du critère).

Nous rejoignons ici l'interprétation de MATALON (1965 - p.25) : "(Dans une échelle hiérarchique), si on sait que tous ceux qui ont résolu un certain problème a en ont également résolu un

autre b, sans que la réciproque soit vraie, on peut penser que les mécanismes intellectuels ou les connaissances qui interviennent dans la résolution de a sont ceux qui sont nécessaires pour celle de b, avec quelque chose en plus. L'unidimensionnalité de la variable sous-jacente prend ainsi une signification précise."

La hiérarchie que nous observons concerne les activités liées au critère par la maximisation de la concordance. Il ne faut cependant pas oublier que l'interprétation dépend essentiellement du critère choisi. La hiérarchie ne représente rien par elle-même : c'est par référence au critère qu'elle est susceptible d'apporter des informations précieuses mettant en évidence des activités souvent difficiles à contrôler.

L'intérêt d'une telle hiérarchie réside dans la cohérence des résultats, qui donne non seulement à l'analyse intrinsèque des activités, mais également au classement des élèves, une validité significative, si l'on a pris soin de bien préciser le critère.

L'étude par l'analyse hiérarchique de classes isolées ou de populations restreintes n'est pas conseillée : le poids de la variabilité individuelle y est trop important. Nous avons, par contre, obtenu des résultats très intéressants dans l'étude de populations plus vastes (à partir de 100 individus). Les effets mis en évidence sont alors de nature à servir l'évaluation informative (annexes 1, 4 et 5).

Nous avons rencontré un intérêt tout particulier pour les structures hiérarchiques dans des domaines très variés.

a) Les techniques destinées à graduer les attitudes

Les travaux de GUTTMAN (1944) sur les données qualitatives n'ont guère été exploités dans la pratique. D'après GREENE (1954), le fait que le modèle de GUTTMAN n'est pas probabiliste constitue une difficulté majeure pour son utilisation. Des études théoriques

ont été effectuées, qui ont précisé certaines caractéristiques du scallogramme. Ainsi STOUFFER et Coll. (1952), cités par GREENE (1954), avaient remarqué l'influence néfaste des items obtenant des pourcentages de réussite extrêmes lors du calcul du coefficient de reproductibilité. LOEVINGER (1948) calcule un coefficient d'homogénéité à partir de la covariance subsistant dans la matrice. EYSENCK et CROWN (1949), cités par GREENE (1954), prétendent qu'un facteur commun, beaucoup plus que la reproductibilité, est la caractéristique essentielle du scallogramme. GREENE (1954) énumère quatre conditions qui doivent être satisfaites pour que le coefficient de reproductibilité mérite d'être mesuré (p. 356) :

1. Une bonne répartition des pourcentages de réussite des items est nécessaire. Cependant, quelques items devraient présenter un pourcentage de réussite inférieur à 20 % ou supérieur à 80 %.
2. Chaque catégorie de réponses devrait avoir plus de "non-erreurs" que d'"erreurs". Sinon, la catégorie ne serait pas corrélée avec l'échelle.
3. Dans le cas où les questions sont dichotomiques, il faudrait au moins utiliser 10 questions.
4. Les erreurs devraient être aléatoires. Si une erreur apparaît systématiquement, c'est la preuve d'un type non hiérarchique.

Ces caractéristiques sont certes intéressantes, mais elles se justifient surtout lorsqu'on souhaite conserver tous les items introduits au départ. GUTTMAN refusait d'ailleurs de rejeter le moindre item, même très pauvre, de manière à conserver le caractère exhaustif de son univers...

Nous pensons que ce qui a surtout nui au développement du scallogramme, c'est la prétention des auteurs de le construire a priori, en créant des questions qui devaient toutes s'intégrer dans une hiérarchie.

Ainsi que nous n'hésiterons pas à le faire, EDWARDS et KILPATRICK (1946) admettent pour leur part la sélection d'items hiérarchisés :

"Toute technique serait de grande valeur si elle nous permettait de sélectionner un groupe d'items parmi un grand nombre d'items possibles, avec une certitude relative que ce groupe sélectionné remplirait, à son tour, les conditions requises pour une analyse de l'échelle "

Les auteurs récents se sont désintéressés du scallogramme au profit de techniques multidimensionnelles.

b) Les taxinomies utilisées dans les sciences de la nature

Dans les sciences de la nature, la taxinomie vise deux objectifs contradictoires :

1. Disposant les individus selon une hiérarchie de classes, elle cherche à décrire l'équilibre des espèces, leurs ressemblances et leurs parentés (démarche descendante). Ainsi, ADANSON (1757), cité par BENZECRI (1973) :
   "Je me contenterai de rapprocher les objets suivant le plus grand nombre de degrés de leurs rapports et de leurs ressemblances... Les objets ainsi réunis formeront plusieurs petites familles que je réunirai encore ensemble, afin d'en faire un tout dont les parties soient réunies et liées intimement."
2. Dans la pratique quotidienne, elle veut déterminer les individus, c'est-à-dire reconnaître leur nom. Tel est l'objet des clefs qui font procéder du général au particulier. L'analyse des caractéristiques présentes ou absentes (items) chez un individu donné permet l'identification de celui-ci (flores dichotomiques, clefs analytiques en chimie, etc.).

Dans nos recherches, c'est le premier de ces objectifs que nous poursuivons.

La taxinomie se fonde sur la notion de distance pour quantifier les différences et les ressemblances observées (BENZECRI - 1973 - p. 67) :

"Un indice de similarité sur un ensemble K est une fonction $S_{kk'} = S_{k'k}$, à la valeur réelle, définie par les paires, k, k', d'éléments de K ; S n'est pas assujettie aux axiomes d'une distance, on demande simplement que S ait une valeur maxima, atteinte seulement pour toute paire formée de deux éléments identiques : $k \in K$, $S_{kk} = v$.
Mais il est possible de définir sur K une vraie distance qui ne soit fonction que de S."

Dans le même ordre d'idées, SOKAL et SNEATH (1963) appellent R - technique la classification selon les caractères (qui est peu étudiée) et Q - technique la classification d'après la similitude des individus. C'est bien entendu dans le cadre de la première technique, curieusement négligée par les chercheurs, que nous nous situons.

c) L'analyse binaire

FAVERGE a appliqué l'analyse binaire à l'étude de certaines attitudes ou opinions :

- l'attitude étudiante au regard de la participation à l'Université (1973);
- les styles de travail des étudiants (1974);
- l'implication d'ouvriers dans leur travail (1975).

A chaque question à libellé bipolaire (par une réponse, le sujet se situe entre deux pôles sur une échelle graduée de 1 à 5), on associe deux colonnes correspondant aux deux sens de cotation de l'échelle, ce qui rend plus claire la représentation et facilite l'interprétation.

Dans une première étape, on entreprend une analyse typologique (1) des sujets. FAVERGE (1973) justifie cette analyse par deux raisons :

1. réduire la grandeur du tableau à analyser, sans cependant perdre une information notable ;
2. fournir une description de la population en termes de personnages types.

---

(1) ANDERBERG Cluster analysis (1973)

L'analyse binaire proprement dite, qui succède à l'analyse typologique, s'apparente à l'analyse des correspondances de BENZECRI. Cependant, BENZECRI étudie les correspondances dans des tableaux relatifs à des dénombrements, alors que l'analyse binaire classique compare des variables.La technique de BENZECRI utilise la distance distributionnelle (ou du XHI^2) et l'analyse binaire se fonde sur la distance quadratique classique.

La méthode proposée vise à obtenir une approximation des distances aussi bonne que possible dans un espace de dimensions aussi faibles que possible. Si une représentation plane est accessible, on voit immédiatement sur le dessin comment les observations se groupent et se distinguent.

d) La typologie psychométrique des tests

CARDINET (1972) accompagne la représentation théorique de la matrice hiérarchique typique d'un certain nombre de considérations sur les caractéristiques techniques qu'elle devrait présenter :

"Dans un cas idéal, on pourrait souhaiter que tous les sujets soient classés sans ex-aequo, de même que toutes les questions. On obtiendrait ainsi du côté des sujets comme des questions des distributions rectangulaires. Les intercorrélations entre les questions (calculées avec des coefficients adaptés à ce cas particulier, comme le H de Loevinger) devraient toutes être élevées."

En ce qui concerne les procédures de contrôle de la hiérarchie, CARDINET (1972) déclare :

"Le modèle hiérarchique étant donné, il faut d'abord examiner s'il rend compte de l'ensemble des réponses et ensuite purifier le test des questions qui ne satisfont pas bien le modèle général. Le calcul d'un coefficient de reproductibilité après reclassement des sujets et des items pour faire apparaître la structure triangulaire des bonnes réponses, peut répondre à la première

question, relative à la pertinence du modèle. Le regroupement ou le rejet des questions déviantes est fait ensuite de façon à améliorer l'accord de l'échelle résultante avec le modèle hiérarchique de base."

e) Conclusions

GUTTMAN et ses successeurs s'orientent, comme CARDINET, vers la mise en évidence d'une structure hiérarchique totale. C'est donc sur la ressemblance formelle de la structure observée avec la représentation du modèle qu'ils focalisent leur attention. Pour notre part, sans que nous négligions les caractéristiques intrinsèques du modèle Guttmanien, la découverte d'une structure hiérarchique centrée sur un CRITERE nous paraît mieux répondre aux exigences d'une analyse débouchant sur une exploitation pédagogique.

Les techniques multidimensionnelles de BENZECRI et de FAVERGE nous ont été d'un grand secours, puisque nous poursuivons des objectifs très semblables aux leurs. Chaque fois que cela était possible, nous avons appliqué ces techniques aux données que nous soumettions à l'évaluation informative (annexes 1, 4 et 5). Sans aboutir à des résultats fondamentalement différents, nous remarquons que ces techniques sont très sensibles au manque de cohérence des données : toutes les variables sont prises en considération et s'influencent l'une l'autre, alors que les techniques d'évaluation informative écartent rapidement les données génératrices d'interaction.

Les méthodes exposées ci-dessus nécessitent un tableau de données absolument complet : à chaque case du tableau doit correspondre une information. Lorsqu'une observation manque, un élève absent pour une partie d'une épreuve par exemple, l'ensemble des données relatives à la ligne (élève) ou à la colonne (item) est perdue. Nous risquons ainsi, dans l'étude de phénomènes évolutifs par des processus longitudinaux, de voir apparaître l'influence de la variable parasite "mortalité" (1).

(1) VANDEVELDE L. Etude critique des dispositifs expérimentaux en pédagogie. Revue des sciences de l'éducation 2-3, 1973 p. 63

Nous évitons cet écueil par des procédés de neutralisation successive :

1. L'élève absent à un des items de la "famille-critère" est exclu de l'analyse.
2. L'élève absent à des items autres que ceux de la "famille-critère" n'est neutralisé que pour les données absentes.

L'influence de la "mortalité" se voit ainsi fortement atténuée.

2°) Les moyens techniques proposés

A. Recherche de la base d'une hiérarchie : la famille-critère

1. L'indice de similarité

Lors de la recherche d'une différenciation par la technique d'analyse d'une classe, nous avons calculé la concordance entre chaque item et la distribution dichotomisée des notes des élèves à l'aide de l'indice de différenciation spécifique (IDS). Cet indice s'obtient en soustrayant la proportion de réussite des élèves inférieurs de la proportion de réussite des élèves supérieurs. Le même principe est utilisé pour calculer la SIMILARITE entre deux items.

Soit les items A et B proposés à 10 élèves :

	1	2	3	4	5	6	7	8	9	10	
A	+	0	0	+	0	+	0	0	0	0	3
B	+	0	+	0	0	+	+	0	0	0	4

a) Ordonnons les élèves en fonction de la réussite à l'item A :

	1	4	6	2	3	5	7	8	9	10
A	+	+	+	0	0	0	0	0	0	0
B	+	0	+	0	+	0	+	0	0	0

Il y a SIMILARITE lorqu'un élève réussit ou échoue simultanément aux deux items : ++ ou OO.
Dans les autres cas, il y a NON SIMILARITE : +O ou O+.

L'indice de similarité entre l'item-critère (A) et l'item-témoin (B), que nous appelons $COIDS_{AB}$, est obtenu en soustrayant la proportion de non similarité chez les élèves ayant échoué à l'item-critère (A) de la proportion de similarité pour les élèves ayant réussi cet item.

$$COIDS_{AB} = \frac{\Sigma(++)AB}{NA} - \frac{\Sigma(O+)AB}{N - NA}$$

où N = nombre total d'élèves ;
NA = nombre d'élèves ayant réussi l'item-critère (A) ;
$\Sigma(++)AB$ = nombre de similarités chez les élèves ayant réussi l'item-critère (A) ;
$\Sigma(O+)AB$ = nombre de non similarités chez les élèves ayant échoué à l'item-critère (A).

Ici, $COIDS_{AB} = \frac{2}{3} - \frac{2}{7} = .38$

b) Ordonnons les élèves en fonction de la réussite à l'item B :

	1	3	6	7	2	4	5	8	9	10
B	+	+	+	+	O	O	O	O	O	O
A	+	O	+	O	O	+	O	O	O	O

$$COIDS_{BA} = \frac{2}{4} - \frac{1}{6} = .33$$

Le choix de l'item-critère pour le calcul du COIDS n'est donc pas indifférent. Nous avons choisi comme critère l'item dont le pourcentage de réussite est le plus proche de 50 %.

Nous verrons plus loin que les résultats obtenus sont ainsi comparables, sous certaines conditions, à ceux du coefficient de corrélation de point (φ) et au H de LOEVINGER.

Si les pourcentages de réussite aux deux items sont égaux, ou s'ils sont également distants de 50 % de réussite, le COIDS est identique, quel que soit l'item-critère choisi.

2. La similarité maximale

L'indice de similarité entre deux items de même pourcentage de réussite varie de +1 à -1. Lorsque les items ont des pourcentages de réussite différents, cet intervalle est diminué.

Nous devons tenir compte de l'amplitude de cette variation dans l'étude de la similarité entre deux items. Comme seules les valeurs positives du COIDS nous intéressent, nous calculons l'indice de similarité maximum ($COIDS_{MAX}$), c'est-à-dire l'indice de similarité entre deux items parfaitement corrélés, eu égard à leur pourcentage de réussite.

Selon que le pourcentage de réussite à l'item-critère est plus ou moins élevé que celui de l'item-témoin, l'algorithme de calcul est différent :

a) % item-critère $\geqslant$ % item-témoin

$$COIDS_{MAX} = \frac{\%\ \text{item-témoin}}{\%\ \text{item-critère}}$$

b) % item-critère $<$ % item-témoin

$$COIDS_{MAX} = \frac{(\%\ \text{item-témoin} - \%\ \text{item-critère})}{(100\ \% - \%\ \text{item-critère})}$$

3. La concordance

Nous appelons CONCORDANCE le rapport $COIDS/COIDS_{MAX}$ entre les deux items. La concordance varie, comme les coefficients de

corrélation, de +1 à -1. Nous avons observé que la concordance présente la particularité d'être égale au rapport φ/φ_{MAX} quand le pourcentage de conformité (PCA) entre les deux items atteint une certaine valeur :

$$PCA = \frac{\Sigma(++) + \Sigma(OO)}{N}$$

Voici, à titre d'exemples, quelques unes des observations :

Item B / Item A		Toutes les valeurs sont multipliées par 100				
		φ/φ_{MAX}	COIDS/COIDS$_{MAX}$	PCA	2.PCA-1	COIDS
216	161	16,73	16,73	56,78	13,55	11,09
75	94					
60	65	14,48	14,48	59,89	19,78	8,46
154	267					
149	35	19,59	38,55	51,83	3,66	17,99
228	134					
127	87	11,35	12,95	54,03	8,06	9,52
164	168					

Conclusion :

$$COIDS/COIDS_{MAX} = \varphi/\varphi_{MAX} \quad \text{lorsque} \quad COIDS \leqslant 2.PCA - 1$$

4. La matrice de similarité

Nous calculons les indices de similarité entre chaque paire d'items d'une épreuve et nous les disposons dans une MATRICE DE SIMILARITE. Pour chaque paire d'items, nous y trouvons le COIDS et le $COIDS_{MAX}$. Les items témoignant d'une réussite maximale ou nulle sont écartés, car ils n'ont de toute manière aucune similarité avec les autres items. Il est impensable d'effectuer tous ces calculs sans recourir à l'ordinateur (annexe 1).

L'exploitation de la matrice de similarité est difficile. Le tableau obtenu est très chargé et peu explicite. En transformant les données numériques en symboles variant selon la valeur de la concordance, on obtient un tableau clair et facile à interpréter : la MATRICE SYMBOLIQUE DE CONCORDANCE.

5. La matrice symbolique de concordance

Dans cette matrice (annexes 1 et 4), nous avons présenté différemment les couples dont les deux items ont un pourcentage de réussite assimilable à 50 % (c'est-à-dire compris entre 23 % et 77 %). Nous avons retenu trois niveaux de concordance :

- Concordance = 1
- Concordance > .70
- Concordance > .50

6. Sélection de "familles-critères"

Les informations contenues dans la matrice symbolique de concordance sur le pourcentage de réussite des items et sur leur concordance constituent deux des éléments de base pour la sélection de "familles-critères".

En analysant le contenu des items des "familles-critères", on vérifie la troisième condition : les propriétés communes des activités sollicitées. Les "familles-critères" sont réunies par tâtonnement. L'algorithme de hiérarchisation des items permet ensuite de vérifier les hypothèses ainsi formulées.

B. La hiérarchisation des items

La suite des opérations s'apparente à la technique de mise en évidence d'une différenciation.

1. Dichotomisation de la "famille-critère"

Nous calculons la note de chaque élève pour l'ensemble des items faisant partie du critère. Les élèves sont répartis en "supérieurs" et "inférieurs" par rapport à cette note.

Rappelons que tout élève qui n'a pas été noté pour un des items de la "famille-critère" est automatiquement exclu du calcul de la hiérarchie.

2. Calcul de la concordance

Nous calculons la concordance ($COIDS/COIDS_{MAX}$) entre chaque item et le critère dichotomisé. Les items dont le pourcentage de réussite est égal à zéro ou au maximum sont éliminés. Les élèves qui ont été notés pour l'ensemble des items-critères ne sont neutralisés que pour les items qui n'ont pas fourni d'observation.

3. Sélection des items hiérarchisés

Les items obtenant avec le critère une concordance supérieure à .46 sont extraits de la matrice et regroupés pour former une hiérarchie. Nous calculons la note de chaque élève à l'ensemble des items hiérarchisés (NH).

4. Interprétation

Le contenu de la hiérarchie - et le comportement des élèves face à celui-ci - est à rapporter au critère sur lequel la hiérarchie se fonde. On obtient des informations à la fois sur le matériel utilisé et sur les individus qui ont subi l'épreuve, à condition de ne pas oublier sur quelles bases la hiérarchie s'est constituée.

3°) Commentaires sur la technique proposée

a) Caractéristiques de l'épreuve hiérarchique

D'après CARDINET (1972), l'épreuve hiérarchique présente une répartition équitable des variations entre les élèves (σ^2_e), les items (σ^2_q) et l'interaction (σ^2_i).

0	0	0	0	0	+	1
0	0	0	0	+	+	2
0	0	0	+	+	+	3
0	0	+	+	+	+	4
0	+	+	+	+	+	5
+	+	+	+	+	+	6
1	2	3	4	5	6	

$$\sigma^2_e = \sigma^2_q = \sigma^2_i$$

Quelques remarques s'imposent :

1. L'égalité des variations entre les élèves et les items n'est possible qu'à deux conditions :
   - la matrice est carrée,
   - la matrice est triangulaire.
2. L'interaction est inévitable. Elle provient de la différence entre les pourcentages de réussite observés, aussi bien chez les élèves que pour les items.
   Toutes les concordances sont maximales :

$$\text{COIDS/COIDS}_{MAX} = \varphi/\varphi_{MAX} = 1$$

3. Nous avons vu que GUTTMAN (1944) mesure la déviation entre la matrice observée et la matrice type par le Coefficient de Reproductibilité (Rep). GUTTMAN fixe à 85 % le seuil de mise en évidence d'un scallogramme. LOEVINGER (1948) critique ce seuil, avec d'autres auteurs, et propose 90 %. Cet indice ne nous est d'aucune utilité, car les erreurs sont calculées

entre les items successifs classés par ordre croissant de réussite. Notre référence à une "famille-critère" ne signifie nullement que les items sont hiérarchisés entre eux : ils ne le sont que par rapport au critère !

4. GUTTMAN avait, lui aussi, remarqué l'existence d'une corrélation maximale entre deux items du scallogramme. Il l'expliquait par la non-normalité de la distribution.

b) L'indice de concordance

L'indice de concordance s'apparente aux coefficients de corrélation. ANDERBERG (1973) énumère les techniques propres à traiter les problèmes de concordance : corrélation simple, régression linéaire, analyse discriminante, analyse factorielle, corrélation canonique, n'importe quelle autre méthode basée sur un modèle linéaire ! Il ajoute (p. 29) :
" Ces techniques sont généralement employées avec des données mesurées sur des échelles d'intervalle. Mais dans le cas de variables dichotomiques, l'échelle de définition n'est d'aucune utilité en raison du fait que seule une échelle nominale est nécessaire. Ainsi, des variables qualitatives dichotomiques peuvent être directement incluses dans de telles analyses fondées sur des échelles d'intervalle."

GOODMAN et KRUSKAL (1954, 1959, 1963, 1973) font l'inventaire des coefficients de corrélation existants. Parmi ceux-ci, celui de PEIRCE qui étudie, en 1884, la prédiction des ouragans.

		OCCURRENCE		
		Tornado	No Tornado	
PREDICTION	T	P_{11} 11	P_{12} 14	$P_{1.}$ 25
	NT	P_{21} 3	P_{22} 906	$P_{2.}$ 909
		$P_{.1}$ 14	$P_{.2}$ 920	$P_{..}$ 934

$$\theta = \frac{P_{11}}{P_{.1}} - \frac{P_{12}}{P_{.2}}$$

θ n'est rien d'autre que notre COIDS, mais en prenant systématiquement la prédiction comme critère. Pour PEIRCE, cet indice résout les difficultés soulevées par le pourcentage élevé de certaines concordances.

YOUDEN (1950) applique le même indice, qu'il appelle J, pour le diagnostic du cancer.

Individuals	Positive	Negative	Total
Know diseased	a	b False négatives	a + b
Healthy on control	c False positives	d	c + d

$$J = \frac{ad - bc}{(a + b)(c + d)}$$

Cette formule est une transformation algébrique du COIDS. YOUDEN estime que son indice est particulièrement apte à détecter les erreurs de diagnostic, c'est-à-dire l'interaction...

L'erreur-standard de J serait calculable à partir de 20 observations :

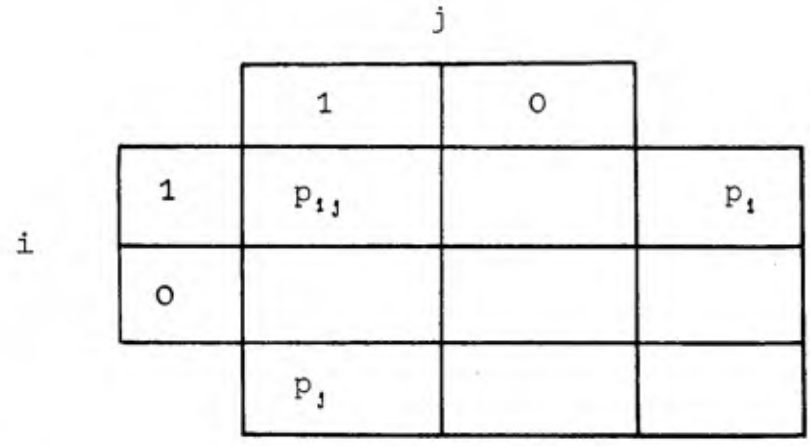
$$S_J = \sqrt{\frac{ab}{(a + b)^3} + \frac{cd}{(c + d)^3}}$$

On pourrait donc comparer les indices de deux populations par un test t de STUDENT.

SOKAL et MICHENER (1958), cités par SOKAL et SNEATH (1963), proposent comme coefficient d'association le pourcentage de conformité (PCA) qui intervient, comme nous l'avons vu, dans la liaison entre la concordance ($COIDS/COIDS_{MAX}$) et le rapport φ/φ_{MAX}.

LOEVINGER (1948) analyse d'une manière détaillée le comportement de la matrice hiérarchique (Cumulative Homogeneous Test). Elle l'assortit d'un Coefficient d'Homogénéité (H) :
" Une moyenne pondérée des probabilités $p_{i/j}$ affectées à chaque paire d'items, adaptée de telle manière que le coefficient égale zéro pour un test parfaitement hétérogène et égale l'unité pour un test parfaitement homogène."

1. Coefficient d'homogénéité entre deux items



$$H_{ij} = \frac{\frac{p_{ij}}{p_j} - p_i}{1 - p_i}$$

avec $p_j \leqslant p_i$

et $\frac{p_{ij}}{p_j} = p_{i/j}$

H_{ij} donne exactement le même résultat que la concordance (COIDS/$COIDS_{MAX}$) ! La formule proposée par LOEVINGER pour calculer le coefficient d'homogénéité a cependant le désavantage de faire appel à des notions statistiques peu accessibles : la probabilité conditionnelle et la probabilité marginale. L'algorithme employé pour obtenir la concordance (COIDS/$COIDS_{MAX}$) repose sur une succession d'étapes logiques, compréhensibles sans référence aux probabilités. Son analogie avec l'algorithme de mise en évidence d'une famille de différenciation le rend également préférable dans le cadre de l'évaluation informative.

LOEVINGER prétend que son coefficient serait particulièrement bien adapté aux variables non-paramétriques. Comme nous, elle n'a effectué qu'une étude empirique, sans se livrer aux vérifications techniques par l'algèbre.

2. Coefficient d'homogénéité d'un test

$$H_t = \frac{\sum_{i=1}^{m-1} \sum_{j=j+1}^{m} p_j \, q_i \, r_{ij} / \max r_{ij}}{\sum_{i=1}^{m-1} \sum_{j=1}^{m} p_j \, q_i}$$

où $r_{ij} = \varphi$

$\max r_{ij}$ = valeur maximale de r_{ij} (φ_{MAX})

Comme le coefficient de reproductibilité de GUTTMAN, le coefficient d'homogénéité d'un test de LOEVINGER s'adresse à l'ensemble de la matrice. Il a cependant l'avantage, sur celui de GUTTMAN, de comparer entre elles toutes les paires d'items sur la base du rapport φ/φ_{MAX} pondéré. La pondération par

la proportion de réussite que LOEVINGER introduit est parfaitement justifiée pour l'étude d'une structure hiérarchique totale, où les pourcentages de réussite extrêmes risquent de fausser considérablement les résultats. Nous avons vu précédemment que la concordance (et H_{ij} par conséquent) n'était comparable au rapport φ/φ_{MAX} que sous certaines conditions. Le résultat obtenu au H_t est donc faussé, puisque LOEVINGER base directement sa formule sur le rapport φ/φ_{MAX} qui ne reflète pas toujours la concordance entre les paires d'items.

Nous avons introduit, pour notre part, un indicateur semblable à celui de LOEVINGER, que nous avons appelé " IDG CRITERE ". Il se calcule en effectuant la moyenne arithmétique des concordances entre chaque item hiérarchisé et le critère. Accompagné des paramètres de statistique descriptive (moyenne, écart-type, indices de PEARSON), il donne des indications utiles sur les caractéristiques de la hiérarchie mise en évidence (annexe 1).

c) <u>La hiérarchie autour d'un critère</u>

La plupart des auteurs qui recherchent des hiérarchies espèrent fonder celles-ci sur des "univers" d'items ou d'élèves en fonction de caractéristiques formelles. LOEVINGER (1948) n'hésite pas à écrire (p. 525) :

" Des items à regrouper peuvent être entièrement déterminés par la relation entre les réponses données aux items choisis uniquement en fonction de similitudes formelles indépendamment de toute hypothèse de la part de l'investigateur."

Mais LOEVINGER écrit quelques lignes plus loin :

" Cependant, il n'est pas recommandable de démarrer la procédure avec des items choisis uniquement en fonction de similitudes formelles indépendamment de toute hypothèse."

Théoriquement, n'importe quelle distribution des élèves conduirait à une hiérarchie. Il suffirait de trouver les items susceptibles de justifier le classement attendu ! Le point de

départ d'une hiérarchie utile pour l'éducateur nous semble mériter davantage de précautions, même s'il est souvent choisi intuitivement ! " Scale analysis methods are intuitive and not rigorous." (LOEVINGER - 1948 - p. 513)

Nous remarquons enfin que les constructeurs d'échelles hiérarchiques accordent une très grande importance à la cohérence du contenu des épreuves.

" La condition la plus importante à laquelle doit répondre un test dont les résultats seront interprétés en termes de mesures semblerait être que tous les items fassent appel à la même catégorie d'aptitudes et d'attitudes." (CURETON, cité par LOEVINGER - 1948).

La cohérence de l'objet traité semble fondamentale, mais nous verrons (annexe 1) qu'il est utile d'introduire des items réclamant des activités différentes pour définir des séries hiérarchisées étalées sur l'ensemble du processus pédagogique, notamment pour la mise en évidence de prérequis.

MATALON (1956) voit une différence entre l'homogénéité d'une hiérarchie, où la répétition des erreurs est étudiée sur les élèves, et la consistance, qui s'attache aux items. Il conseille de limiter l'échelle aux items dont le pourcentage de réussite est compris entre 10 % et 90 %. L'emploi de la concordance autorise l'analyse des items extrêmes, bien que la prudence s'impose lors de leur interprétation.

Le choix d'"items critères", "pivots" autour desquels s'articule la hiérarchie que nous recherchons, nous éloigne des conceptions de GUTTMAN et de ses successeurs. Notre objectif est fort différent du leur : nous ne cherchons pas à déterminer des niveaux d'attitudes chez des individus, mais à découvrir la nature des effets didactiques qui conduisent à des observations hiérarchisées.

d) Les étapes de la recherche d'une hiérarchie

Notre technique rejoint les propositions de MATALON (1965), qui définit les étapes de la vérification d'une hiérarchie comme suit :

1. Choix d'une échelle parfaite qu'on suppose "proche" des données à analyser, c'est-à-dire que, pour chaque rang, on définit un patron-type (pattern) appelé à servir de référence.
2. Attribution d'un rang à chaque patron effectivement rencontré dans l'échantillon, donc à chaque sujet.
3. Comptage, par sujet, du nombre de réponses qui diffèrent des réponses du patron-type : les erreurs.
4. Introduction du nombre d'erreurs dans un coefficient, pour lequel on a choisi une valeur critique.

Pour notre part, nous nous sommes tourné vers la détermination de "familles-critères" plutôt que vers la mise en évidence d'une hiérarchie totale orientée vers les sujets.

D'après MATALON, les techniques diffèrent selon :

- le choix de l'échelle de référence ;
- les rangs attribués aux sujets ;
- le coefficient choisi.

Les techniques accordent une plus ou moins grande importance aux tâtonnements.

MATALON conclut avec pessimisme (p. 57) :

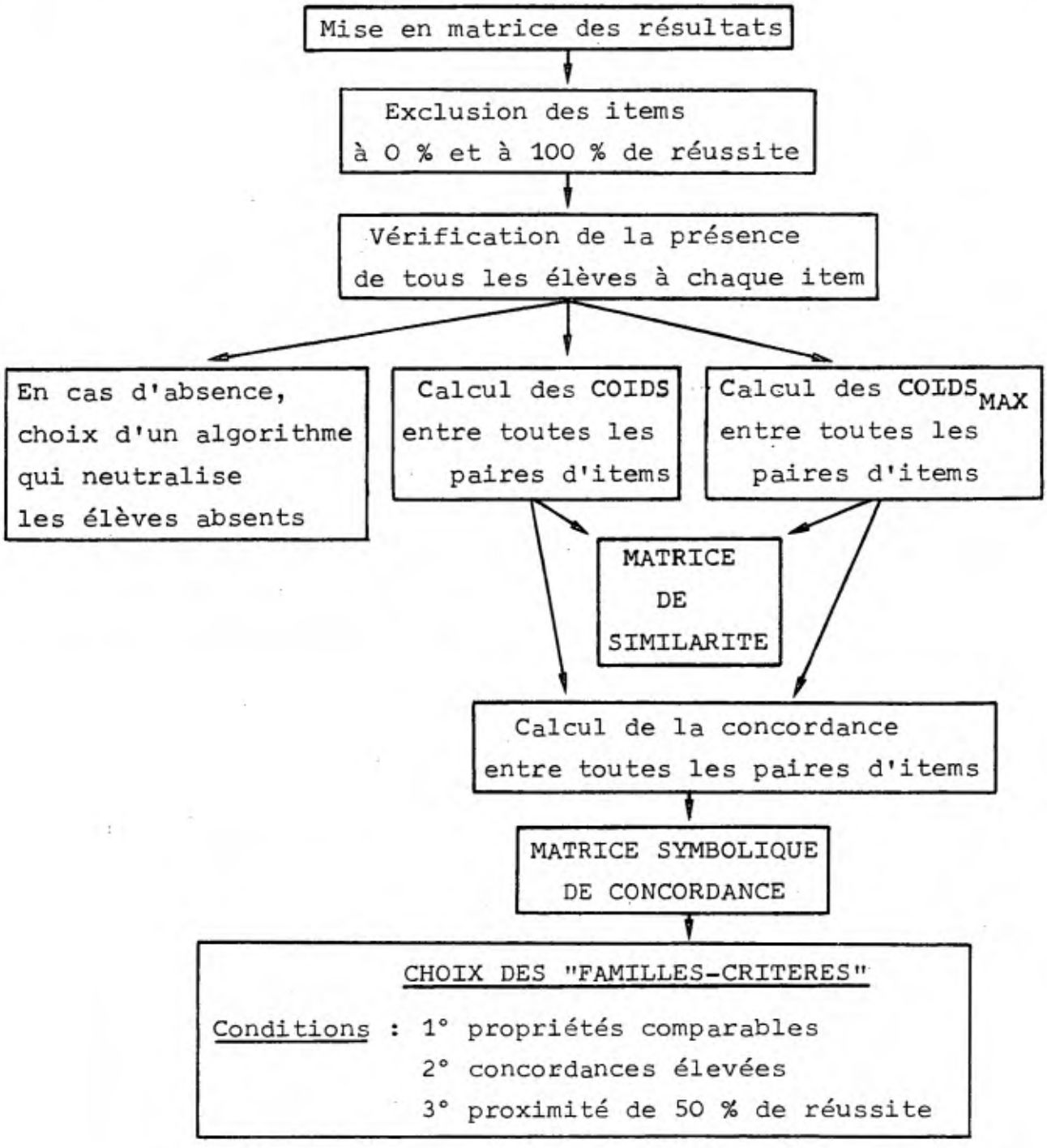
" Il n'existe actuellement aucune technique générale pour résoudre le problème de la construction d'échelles cumulatrices. Tout ce qu'on peut faire, c'est partir d'un nombre suffisamment grand d'items choisis plus ou moins arbitrairement et dont on suppose, pour des raisons quelconques, qu'ils pourront constituer une bonne échelle, utile pour le but poursuivi."

Par l'élaboration d'une hiérarchie axée sur un critère, nous répondons à une des principales critiques que l'on formulait à l'égard du scallogramme : toute notre attention est

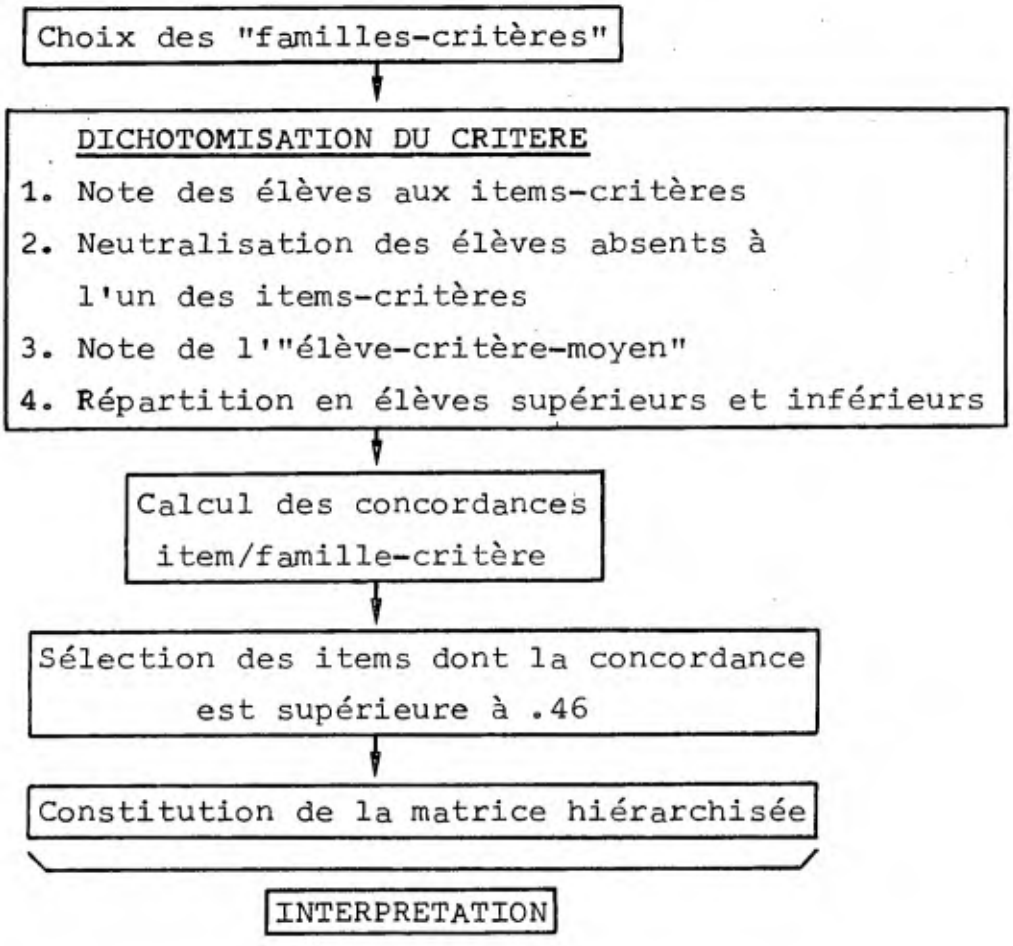
centrée, non pas sur la recherche d'une structure formelle théorique, mais bien sur l'interprétation et l'exploitation d'informations de nature didactique.

4°) Tableaux schématiques de la technique

a) Recherche de "familles-critères"



b) Hiérarchisation des items



5°) Conclusions

Les hiérarchies nous fournissent des informations étalées sur l'ensemble des items, quel que soit leur pourcentage de réussite. Notre souci de rechercher des résultats cohérents, reflétant des activités pédagogiques, nous a conduit à articuler les hiérarchies autour de critères bien délimités, qui définissent avec précision un champ d'interprétation.

L'exploitation des techniques de mise en évidence de hiérarchies n'est envisageable que dans le cadre d'étude scientifiques, mais les mécanismes utilisés restent accessibles à tous les éducateurs. Nous proposons plusieurs exemples dans les annexes 1, 4 et 5.

III INFLUENCE DES INSTITUTIONS SCOLAIRES SUR LES EFFETS COLLECTIFS

1°) Présentation du problème : Analyse d'une population répartie en classes

L'évaluation informative que nous avons développée jusqu'à présent a pour objet l'étude des effets de l'action éducative sur les individus. Elle fournit des informations, d'une part au professeur, dans le cadre limité de sa classe ; d'autre part aux responsables de l'enseignement, sans que l'appartenance des élèves à des classes et à des institutions particulières soit prise en considération.

L'entité "classe" a une influence capitale sur l'activité pédagogique. En outre, pour des raisons pratiques et déontologiques, " l'unité observée en pédagogie est souvent la classe, c'est-à-dire l'ensemble professeur-élèves réunis." (HALLEUX et VANDEVELDE - 1972). Institutionnellement, la stratification par niveaux et l'uniformisation des programmes tendent à mettre sur un pied d'égale progression tous les élèves confrontés à une même matière. L'effet collectif n'est donc pas seulement attendu sur les individus pris isolément, mais tout est mis en oeuvre pour le faire apparaître simultanément dans toutes les entités pédagogiques.

De même qu'il existe des différences entre les individus ayant bénéficié d'une action pédagogique commune, de même nous observons entre les classes d'un niveau donné des discordances qui proviennent :

-soit des conditions didactiques (influence de l'enseignant);
-soit de caractéristiques locales (notamment la taille de l'établissement);
-soit de la manière dont les classes ont été constituées (classes homogènes ou parallèles);
-soit des circonstances dans lesquelles la prise de mesure a été effectuée (effet de contexte).

Un effet collectif est constaté lorsque toutes les classes d'un même niveau réagissent de la même manière aux activités qui leur sont proposées. Nous disons alors que ces activités sont STABLES, c'est-à-dire que toutes les classes se comportent à leur égard d'une manière homogène. En d'autres termes, les activités qui paraissent sensibles à certaines variations aléatoires - notamment l'expérience antérieure - et qui donnent lieu à des observations dissemblables selon les classes, ne relèvent pas d'un effet collectif.

Lorsque les conclusions sont destinées à être généralisées, on prendra toutes les précautions expérimentales nécessaires, notamment en matière d'échantillonnage. Par contre, si la validité interne suffit - ce qui est souvent le cas lors de l'analyse de l'impact d'une didactique - il n'est pas indispensable de recourir à des dispositifs compliqués.

Nos investigations se font dans deux directions :

a) la stabilité dans des activités isolées, où chaque item est étudié séparément (technique "ARC SINUS");
b) la stabilité dans des activités complexes, qui sont suscitées par des épreuves complètes ou des groupes d'items (technique XHI^2).

2°) Moyens techniques proposés

1. Stabilité de la population pour un item isolé

La technique "ARC SINUS" (DAGNELIE - 1970) détermine un intervalle de confiance autour d'une proportion donnée. Lorsque les élèves, répartis en classes, ont répondu à un item particulier, on calcule la proportion de réussite sur l'ensemble des élèves testés. La technique "ARC SINUS" vérifie si les proportions de réussite observées dans chacune des classes sont non significativement différentes de la proportion de réussite de l'ensemble de la population. La proportion de réussite totale est appelée x.

$$x = \frac{\Sigma +}{N}$$

où $\Sigma +$ = nombre de bonnes réponses pour l'ensemble des élèves testés;

N = nombre d'élèves testés.

a) Transformation de la proportion de réussite totale

$$y = 2 \arcsin \sqrt{x}$$

où y est la variable transformée;

x est la proportion de réussite totale.

b) Limites angulaires de l'intervalle de confiance

$$\begin{matrix} y_a \\ y_b \end{matrix} = y \pm \frac{U_{1 - \alpha/2}}{\sqrt{\frac{N}{ng}}}$$

où α est le niveau de signification conventionnellement fixé à .05 en pédagogie ($U_{1 - \alpha/2} = 1{,}96$);

N est le nombre total d'élèves;

ng est le nombre de classes.

c) Proportions marquant les limites de l'intervalle de confiance

$$x_a = \sin^2 \left(\frac{y_a}{2} \right)$$

$$x_b = \sin^2 \left(\frac{y_b}{2} \right)$$

x_a et x_b sont les proportions limites.

La technique "ARC SINUS" a été préférée aux autres techniques, particulièrement à la méthode de l'erreur-standard qui utilise directement la loi normale, parce qu'elle produit un intervalle dissymétrique lorsque la proportion de réussite totale étudiée s'éloigne de 1/2. Elle présente l'inconvénient d'être sensible à l'inégalité des effectifs des groupes comparés. Bien que les effectifs des classes soient rarement identiques, ils restent toutefois suffisamment voisins pour que la technique soit appliquée sans danger, avec référence à l'effectif moyen.

Les classes dont les proportions de réussite se trouvent à l'extérieur de l'intervalle de confiance sont éliminées. Les classes qui subsistent sont homogènes par rapport à l'item considéré.

L'exploitation manuelle de la technique "ARC SINUS" est imaginable, si l'on recourt soit à une table de conversion, soit à une calculatrice pourvue de la fonction "arc sinus". Cependant, le travail devient vite fastidieux, car la technique exige parfois plusieurs applications pour chaque item avant d'éliminer toutes les classes qui se situent à l'extérieur de l'intervalle. Le recours à l'ordinateur devient indispensable lorsqu'on veut comparer plusieurs populations et faire varier la composition du groupe total.

Sont instables les items pour lesquels l'ensemble des classes ne se comporte pas de manière homogène, c'est-à-dire ceux qui provoquent un grand nombre de rejets injustifiables. Les items qui présentent un pourcentage de réussite extrême n'apportent par cette technique aucune information exploitable.

Outre qu'elle permet l'analyse pure et simple d'une population scolaire (annexe 1), la technique "ARC SINUS" s'applique encore pour la constitution de sous-groupes équivalents, ou pour l'observation comparée de groupes soumis à des procédures didactiques différentes. Elle apporte également des informations lors de la confrontation de résultats avant et après une expérimentation (annexe 5).

2. Stabilité de la population pour un groupe d'items

HALLEUX (1973) a utilisé certaines propriétés de la distribution XHI^2 pour déterminer l'appartenance d'échantillons d'élèves à une population parent. En appliquant le test XHI^2 à des échantillons variés, elle a tiré des conclusions importantes à propos des techniques d'échantillonnage (p. 114) :
" Si nous prélevons plusieurs échantillons aléatoires simples, la moyenne des moyennes obtenues est le meilleur estimateur de la tendance centrale, alors que $s/\sqrt{n}$ est l'estimateur le plus fidèle de la dispersion."

En considérant que la répartition des élèves en classes s'effectue de manière aléatoire, nous devons observer des différences non-significatives entre la moyenne de chaque classe et la moyenne des moyennes à une épreuve.

La variable $\left[\frac{(\overline{X} - \overline{x})}{\frac{s}{\sqrt{n}}}\right]$

où $\overline{X}$ est la moyenne des moyennes;
$\overline{x}$, s et n sont respectivement la moyenne, l'écart-type et l'effectif de la classe,

$\left[\frac{(X - x)}{\frac{s}{\sqrt{n}}}\right]^2$ se distribue comme $F_{(1,n-1)}$.

Lorsque n est suffisamment grand, la distribution de cette nouvelle variable se rapproche de celle de XHI^2 :
" La propriété d'additivité des XHI^2 est fort importante et est susceptible de nombreuses applications : en particulier si on réplique une expérience avec des individus différents, on peut appliquer le XHI^2 à chaque échantillon mais aussi à l'échantillon global : pour obtenir ce dernier, il suffit d'additionner les XHI^2 dans chaque échantillon; le nombre de degrés de liberté sur le XHI^2 global est égal à la somme des degrés de liberté sur chaque XHI^2." (D'HAINAUT - 1975 - p. 251)

Dans les classes de nos régions, les effectifs n'atteignent pas souvent les 30 élèves que D'HAINAUT juge nécessaires pour une utilisation parfaite de la technique. Nous n'en sommes cependant pas éloignés, et dans les pays en voie de développement, le nombre d'élèves par classe dépasse toujours cette limite.

Les variables XHI^2, en s'additionnant, donnent une vue d'ensemble de la distribution des classes. Pour le groupe d'items testés, on sélectionne les classes qui appartiennent à une même population, donc celles pour lesquelles on constate un effet éducatif concordant, et l'on élimine les classes qui conduisent à un rejet de la somme des valeurs XHI^2.

Appliqué à un échantillon occasionnel, le test XHI^2 retient les classes qui s'écartent au minimum de la distribution de la population parent visée. En effet, dans l'enseignement et pour des raisons déontologiques, seuls les échantillons occasionnels ou empiriques sont pratiquement accessibles.

En multipliant les observations et en faisant varier les situations, nous avons trouvé dans cette technique un précieux auxiliaire pour étudier l'homogénéité d'un groupe de classes (annexes 1 et 5). HALLEUX (1973) n'avait pu envisager que l'analyse d'épreuves complètes. L'ordinateur sélectionne facilement des subtests ou des groupes d'items pour étudier le comportement de la population à leur égard.

Nous avions espéré trouver des liaisons entre les analyses effectuées par la technique centrée sur l'item (ARC SINUS) et la technique globalisante (XHI^2). Les résultats sont peu encourageants. L'impact de la variable "classe" est trop important pour autoriser une synthèse des deux approches.

3°) Commentaires sur les techniques proposées

a) La technique "ARC SINUS" conduit à une analyse très fine du comportement des items étudiés. Le pourcentage de réussite

rectifié, après élimination des classes hétérogènes, rend mieux compte de la difficulté réelle de l'item.

Cependant, cette technique néglige deux effets importants :

1. L'effet de contexte

Il n'est pas tenu compte des circonstances qui entourent la passation de l'épreuve, et spécialement de la position d'un item par rapport aux autres.

2. L'apport de chaque individu

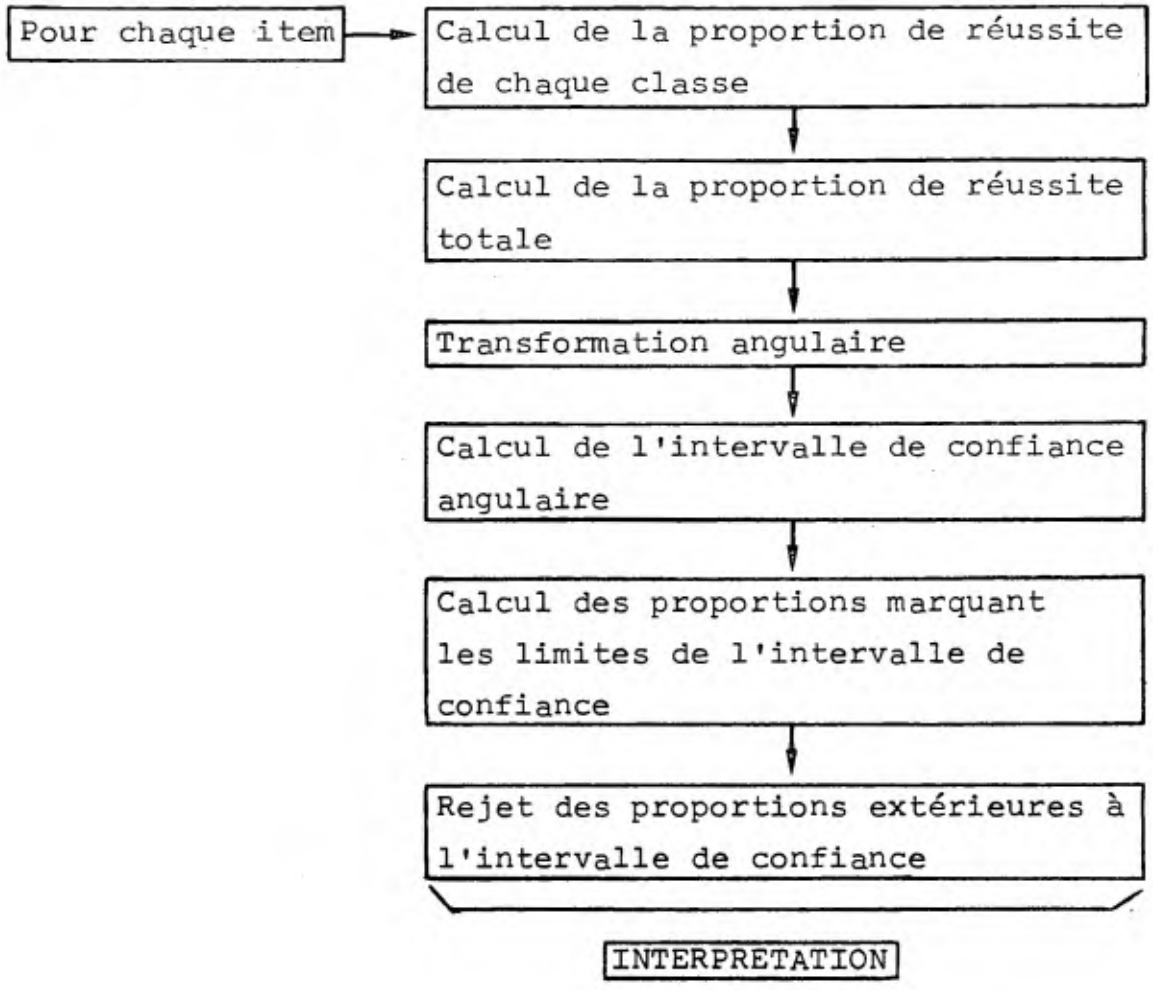
Seul intervient le pourcentage global enregistré dans chaque classe. Le regroupement d'items homogènes en familles est rendu précaire par des variations individuelles incontrôlées.

b) La technique XHI^2 compare le comportement de classes confrontées à un groupe d'items. Les individus des classes étudiées interviennent dans le calcul par l'intermédiaire de l'écart-type. Une classe parfaitement homogène, donc à l'écart-type nul, serait automatiquement exclue... Par contre, lorsque le groupe d'items constitue une famille de différenciation, l'écart-type atteint son maximum et augmente les chances de voir la classe admise dans l'ensemble de la population.

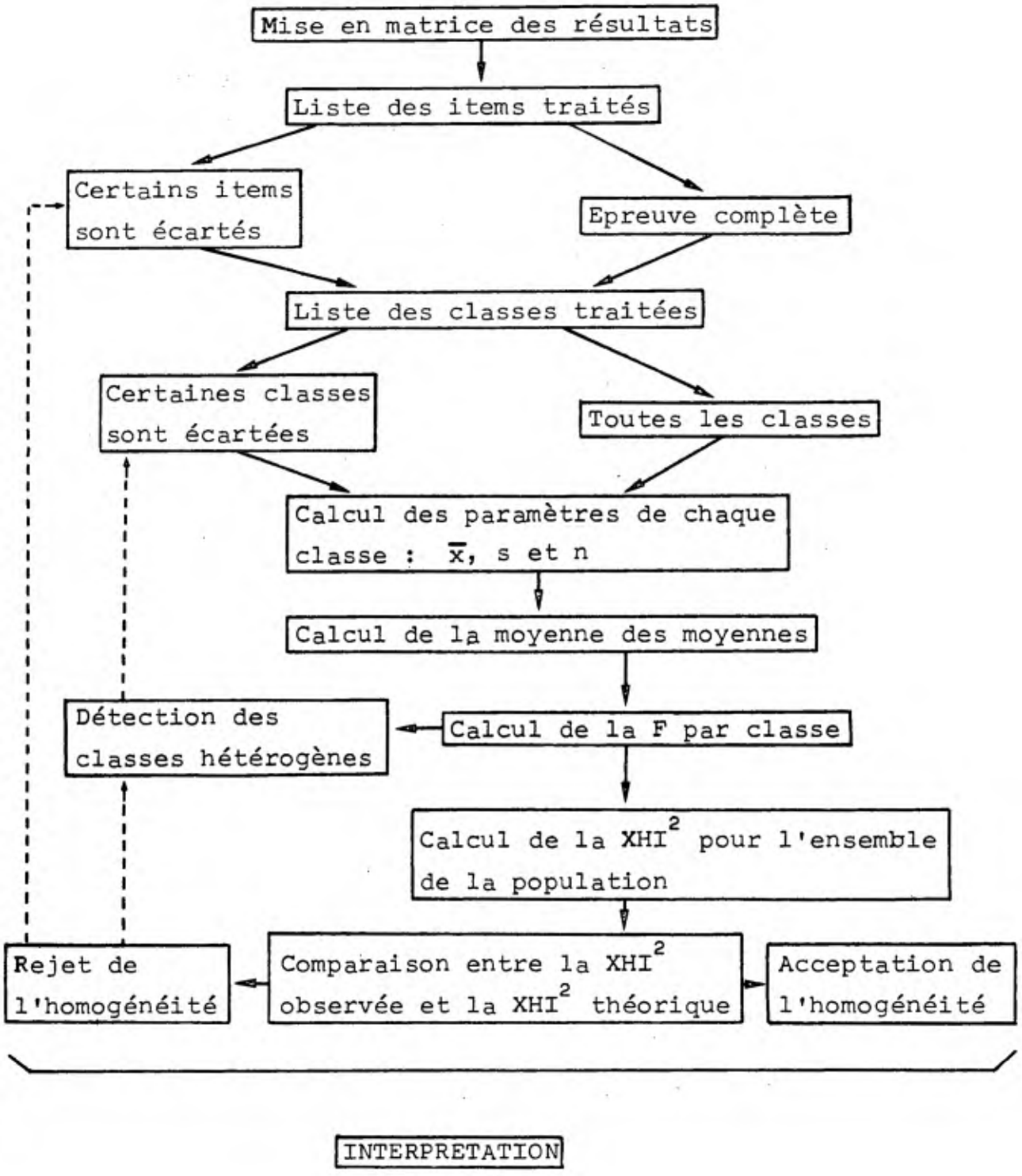
Nous avons eu l'occasion d'étudier le comportement de nombreuses classes dans des circonstances très variées. Nous avons notamment observé des classes de constitution homogène (regroupant les "meilleurs" élèves), nettement supérieures à la moyenne dans des épreuves de connaissance et d'utilisation de l'acquis, mais qui rentraient dans le rang pour des épreuves de compréhension ou d'analyse. Les critères qui avaient présidé à la sélection des élèves se trouvaient ainsi mis en évidence (annexe 1).

4°) Tableaux schématiques des techniques

a) Technique "ARC SINUS"



b) Technique "XHI^2"



5°) Conclusions

Le cadre dans lequel l'activité éducative se déroule a une influence directe sur les résultats enregistrés. Les élèves étant répartis dans des entités pédagogiques distinctes (les classes), les méthodes d'évaluation informatives visent également la mise en évidence des effets collectifs produits par l'institution.

La stabilité des activités évaluées et l'homogénéité de la population rendent compte d'une progression régulière vers les objectifs prévus.

CHAPITRE IV MISE A L'EPREUVE DES METHODES D'EVALUATION INFORMATIVE

INTRODUCTION

Les méthodes d'évaluation informative que nous venons d'exposer ne sont pas simplement spéculatives. Elles ont été en permanence confrontées à la réalité des faits, d'abord pour leur mise au point, ensuite pour la vérification de leur fonctionnement et de leur utilité.

La production de tests standardisés n'entrait nullement dans nos objectifs. Nous n'avions donc pas à construire les instruments de mesure générateurs de données. Nous n'avions pas non plus à participer aux prises de décision consécutives à nos interventions, d'autant moins que l'interprétation des résultats reste liée à des circonstances extérieures, à propos desquelles nous ne recevions la plupart du temps que peu d'informations. Nous ne pouvions nous fixer aucun OBJET, puisque notre étude portait sur une **METHODE**.

Nous avons sélectionné un certain nombre de situations pour l'emploi de nos techniques. Nous avons réuni un ensemble de circonstances des plus variées pour que la conception et la réalisation de nos méthodes s'appuient sur des bases aussi solides que possible. De cette manière, nous avons obtenu une vue générale de la question. Une utilisation plus fructueuse encore de nos techniques nous paraît envisageable dans un avenir proche, à condition qu'un minimum de moyens soit affecté à leur diffusion et à leur exploitation. Pour ne pas alourdir notre texte, nous présentons l'essentiel de ces études en annexe. En voici cependant un bref résumé.

ANNEXE 1 Recherche systématique de longue durée en collaboration avec des enseignants de mathématique

Les travaux que nous avons réalisés avec une équipe de professeurs de mathématique de la région de Marche-en-Famenne

nous ont apporté énormément de matériel d'évaluation. C'est souvent grâce à la collaboration de ces professeurs et à leurs remarques judicieuses que nous avons réuni les conditions nécessaires à la mise au point des techniques d'évaluation informative.

Nous présentons l'historique de nos interventions et les programmes mécanographiques que nous avons été amené à rédiger pour traiter les données recueillies. Parmi les épreuves qui ont été construites et exploitées, nous avons dû effectuer une sélection. Pour chacune des deux années d'études visées, nous n'avons développé en détail qu'une seule épreuve. Nous y avons déployé toutes les techniques applicables au matériel utilisé, et nous avons effectué des comparaisons avec les résultats enregistrés par l'analyse des correspondances sur ces mêmes données.

ANNEXE 2 Application des techniques d'évaluation informative à la construction d'épreuves standardisées

Dans le cadre de recherches expérimentales en vue de construire des tests standardisés pour le cycle d'orientation de l'enseignement secondaire rénové, les techniques d'évaluation informative ont été appliquées sur les résultats des épreuves préparatoires. Les informations recueillies ont été exploitées pour la constitution de tests diagnostiques et pronostiques.

ANNEXE 3 Présentation simplifiée de l'évaluation informative dans le cadre de la formation des enseignants

Si les principes que nous avons développés dans les chapitres précédents nous semblent accessibles à tous les enseignants, même à ceux qui n'ont reçu aucune formation mathématique spéciale, la présentation de la technique de base s'accompagne d'explications complémentaires et recourt à des détours didactiques.

Nous nous sommes fondé sur une démonstration soulignant l'ambiguïté des techniques traditionnelles d'évaluation de l'élève, et nous avons proposé en contraste les principes de l'évaluation informative dans le cadre d'une classe. Le lecteur trouve également l'analyse détaillée de plusieurs matrices, des exercices et des exemples d'interprétation de résultats.

ANNEXE 4 Etude comparative de données expérimentales selon différentes méthodes d'analyse

Les résultats enregistrés sur un grand nombre d'étudiants de première candidature à une épreuve de compréhension nous ont fourni la matière première pour comparer les techniques d'évaluation informative à l'analyse des correspondances de BENZECRI et à l'analyse binaire de FAVERGE. Nous observons que les résultats de l'évaluation informative sont très voisins, et en tout cas jamais en contradiction avec les résultats obtenus par ces autres techniques.

ANNEXE 5 Utilisation des méthodes d'évaluation informative dans des travaux expérimentaux

Plusieurs chercheurs ont utilisé nos techniques dans leurs travaux expérimentaux. Nous présentons succintement leurs objectifs, les développements auxquels ils ont abouti et des compléments que nous avons jugé bon d'y apporter pour traiter certains aspects de leurs résultats que l'évaluation informative mettait particulièrement en relief.

Nous nous sommes intéressé aux travaux de BAMBA sur l'importance de l'emploi d'une langue vernaculaire d'enseignement au Zaïre, à ceux de FRYDMAN sur le développement de l'aptitude à exploiter un matériel de référence, et enfin aux recherches d'AMEGAN sur l'entraînement à des comportements créatifs.

CONCLUSIONS

La majorité des systèmes d'évaluation scolaire aboutissent à des classements. Nous ne contestons pas l'intérêt du classement des élèves, mais bien le manque de maîtrise du contenu des épreuves qui y conduisent, l'absence de référence aux objectifs éducatifs et aux stratégies didactiques mises en place pour les atteindre.

Les données enregistrées sont additionnées, alors qu'elles traduisent à la fois les effets de compétences individuelles, de circonstances fortuites et d'une action intentionnelle exercée par l'éducateur.

En vue d'une adaptation constante de l'institution scolaire aux élèves, il est indispensable de déceler la part des effets attribuables à l'action éducative proprement dite par rapport à celle des autres sources de modifications comportementales. Le problème revient à chercher des estimateurs valables.

1. L'évaluation informative s'oriente vers l'analyse systématique des données prélevées par l'évaluateur. Celui-ci est incapable de distinguer, au premier abord, les effets réellement imputables à l'action collective entreprise par l'éducateur des influences extérieures à l'institution. L'évaluation informative a pour ambition de déterminer, au travers des informations fournies par le matériel d'évaluation, la part qui provient effectivement de l'action éducative.

2. A condition de reposer sur des contenus valides, les résultats voisins de 100 % et de 0 % relèvent indiscutablement de l'action collective. En effet, tous les individus sont différents les uns des autres ; invoquer les aptitudes individuelles

dans l'explication de la tendance extrême des résultats n'est donc pas logique. Dès lors, l'éducateur s'interrogera sur la pertinence de ses objectifs et sur les moyens qu'il a mis en oeuvre pour les atteindre.

3. L'individualisation de l'enseignement paraît irréalisable dans nos systèmes collectifs. Ne serait-il pas possible, toutefois, d'adapter l'éducation à des sous-groupes circonscrits au sein de la collectivité ? Si des caractères communs sont définissables, le problème consiste à découvrir quel groupe d'élèves tire effectivement profit de la didactique adoptée et quels élèves, au contraire, réclament une intervention différente.

4. Des activités de niveau de difficulté très différent peuvent, elles-aussi, être génératrices d'informations précieuses, à condition que la cohérence des résultats permette d'éliminer l'influence de facteurs aléatoires produisant une interaction incontrôlée. Il en résulte une double hiérarchisation des élèves et des items, qui n'a de sens, dans l'étude des résultats de l'action pédagogique, que lorsqu'elle se fonde sur des hypothèses solides et exploitables autour desquelles la hiérarchie s'organise.

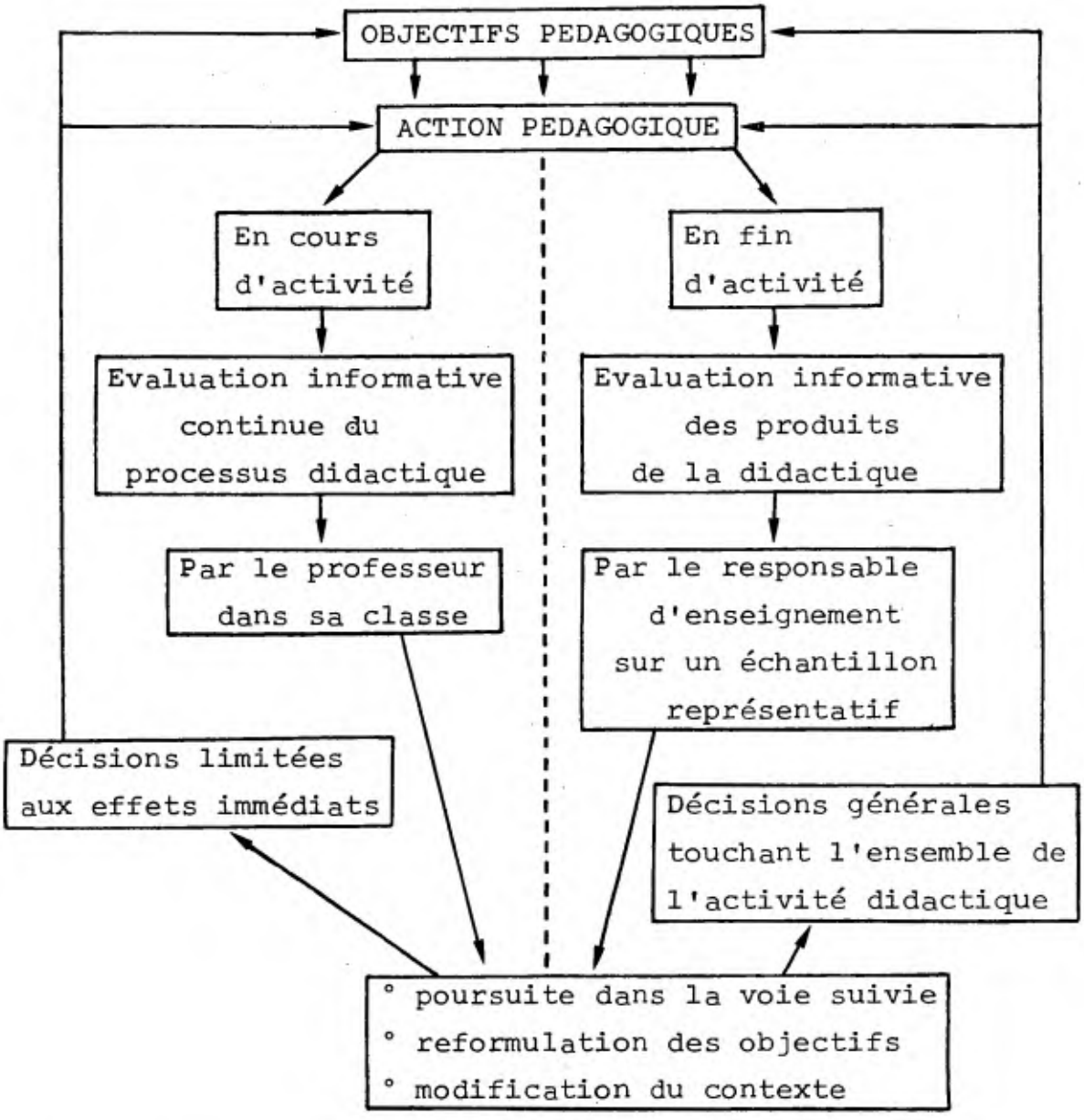
5. Les observations individuelles ne sont pas les seules à présenter un intérêt pédagogique. La classe, élément de base de l'activité didactique, mérite toute notre attention. Les informations recueillies sur les effets qu'y provoque l'action éducative sont utilisables aussi bien dans le cadre d'études expérimentales que pour une analyse interne sans ambition de généralisation. En examinant l'homogénéité des classes testées, on peut s'assurer de l'harmonie des effets de l'action éducative entreprise.

6. Les techniques mises en place pour traiter les mesures effectuées en vue de l'évaluation informative ne se justifient que si elles s'accompagnent de l'interprétation des résultats enregistrés. En effet, l'interprétation constitue, en vue de la décision finale, une synthèse de toutes les informations utiles :

- les objectifs assignés à l'action éducative ;
- les conditions d'application de la didactique ;
- les caractéristiques de la population bénéficiaire ;
- la nature des déclencheurs et des révélateurs de l'évaluation ;
- les circonstances de la prise de mesure ;
- les critères de sélection des items ;
- les techniques utilisées pour le traitement des données.

7. L'évaluation informative s'adresse aussi bien à l'élément de base du système éducatif, le professeur, qu'aux organisateurs nationaux et internationaux. Appliquée au niveau de la classe, l'évaluation informative analyse la manière dont les difficultés présentées aux élèves ont été franchies. L'évaluation concerne donc les tâches elles-mêmes et non les performances des individus. Lorsqu'elle s'élève à l'appréciation de systèmes d'enseignement, l'évaluation informative est destinée à alimenter la critique des programmes, des méthodes, des conditions de travail, des structures institutionnelles, par opposition à une évaluation centrée sur la mise en évidence des aptitudes spécifiques des élèves.

Schéma général de l'évaluation informative

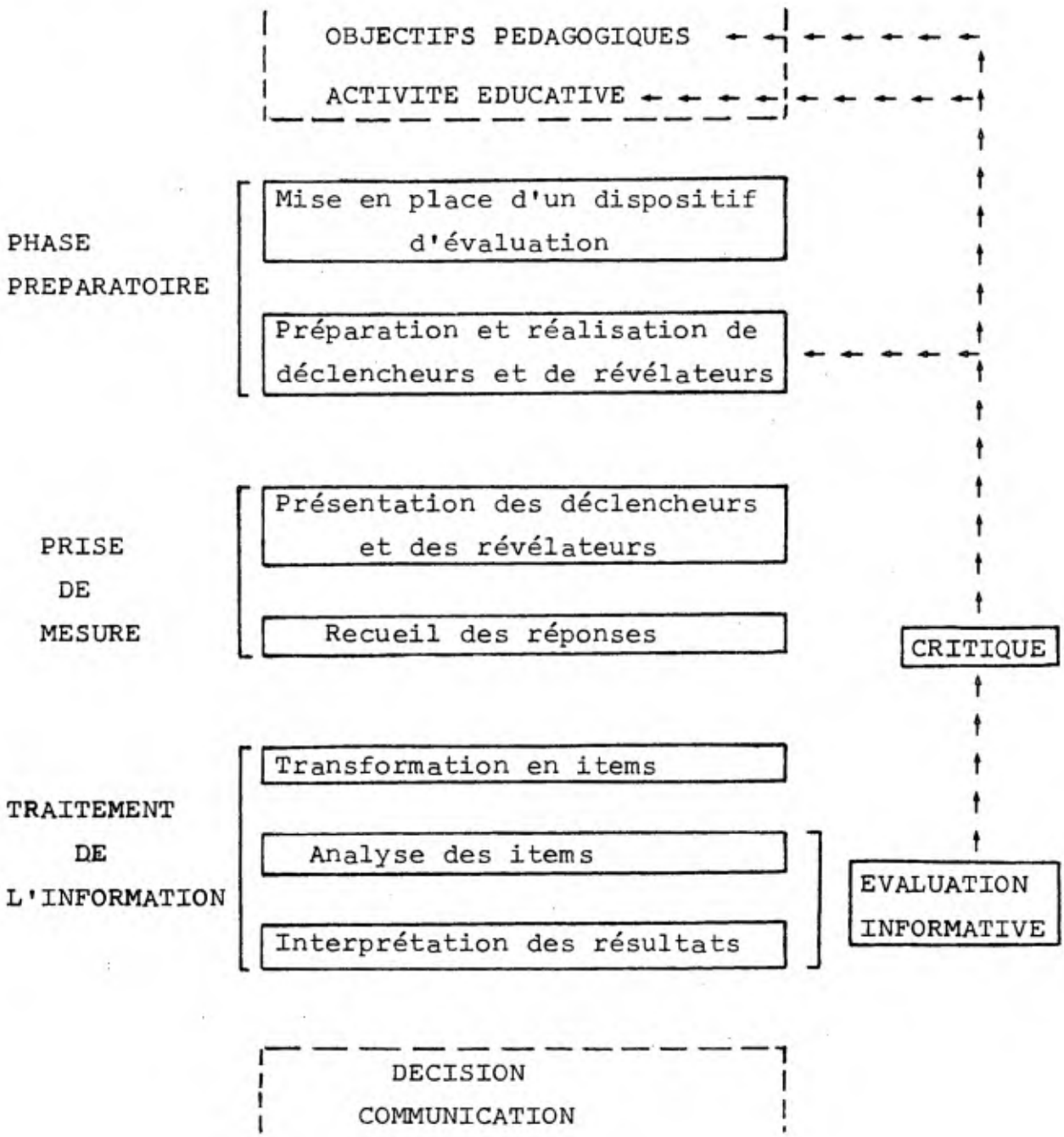


a) Au cours de l'activité pédagogique, l'éducateur prélève des mesures destinées à lui fournir des informations pour l'ajustement de sa didactique. Afin de garder toute leur efficacité, les décisions sont immédiatement appliquées, mais leur ampleur est forcément limitée

b) Après qu'un curriculum a été appliqué, les responsables de l'enseignement interrogent les résultats, afin de prendre des décisions sur les suites à donner à l'action entreprise.



8. Quel que soit le niveau où il se place, le processus d'évaluation assure la liaison entre les objectifs pédagogiques et les décisions qui déterminent la didactique.
Le processus d'évaluation se décompose en trois phases :



a) La phase préparatoire voit la mise en place du matériel d'évaluation.

b) Une phase intermédiaire permet de recueillir les données de base qui fourniront les informations.

c) La phase finale de traitement de l'information comprend l'analyse des données et l'interprétation des résultats.

9. Les résultats de l'évaluation informative ne sont pas destinés à être communiqués tels quels aux élèves. Les renseignements recueillis sont destinés aux éducateurs et aux responsables de l'enseignement, qui souhaitent connaître l'impact de l'action entreprise. Le comportement des élèves n'est intéressant que dans la mesure où il renseigne sur l'efficacité de la didactique choisie.

Cependant, l'élève n'est pas oublié, et son évaluation ne doit pas être sous-estimée. Il apparaît en filigrane à chaque palier de l'évaluation. On le tient régulièrement au fait du résultat de son travail, mais cette communication a pour objectif essentiel de renforcer sa coopération dans le processus pédagogique.

BIBLIOGRAPHIE DE TRAVAIL

ANDERBERG M. R. Cluster analysis for applications.
Academic Press - New York - 1973 - 369 p.

BAMBA-KAMALANDUA Zikadia Le recours à une langue vernaculaire comme moyen d'enseignement à l'Ecole Primaire au Zaïre. Enquête expérimentale developpée en première année primaire. Thèse non publiée - Novembre 1976 - 316 p. + Annexes

BARR, DAVIS, JOHNSON Educational Research and Appraisal
Lippincott - 1953 - 362 p.

BENZECRI J.P. et Collaborateurs L'Analyse des données
1. La Taxinomie.
2. L'Analyse des correspondances.
DUNOD - 1973 - 2 volumes

BERGAN S. B. & DUNN J. A. Psychology and Education : A science for instruction.
John Wiley & sons - New York - 1976 - 542 p.

BLOCH J. R. Quality and quantity in the analysis of integrated science curricula.
Studies in Educational Evaluation - Vol 2 n°2 - 1976 - pp 63 - 83

BLOCK J. H. Mastery learning . Theory and Practice.
Holt, Rinehart and Winston - New York - 1971 - 152 p.

BLOOM B. S. Taxonomy of Educational Objectives.
Handbook I : Cognitive Domain.
David McKay - New York - 1956

BLOOM B. S., HASTINGS J. T., MADAUS G. F. Handbook on Formative and Summative evaluation of Student Learning,
McGraw-Hill Book Company - New York - 1971 - 923 p.

BONBOIR Anna La Docimologie.
PUF - Paris - 1972 - 195 p.

BOURDIEU P. et PASSERON J.C. La Reproduction,
Ed. de Minuit - Paris - 1970 - 282 p.

CARDINET Jean Dépasser la sélection scolaire.
GRETI - Collection Information - Lausanne 1969 - 19p.

CARDINET Jean L'adaptation des tests aux finalités de l'évaluation.
Revue de l'Ecole nouvelle - n° 2-3 - Didier - Paris 1973

CARDINET Jean et TOURNEUR Yvan Une théorie des tests pédagogiques.
IRDP - Neuchâtel - 1974 - 33p.

CARDINET J., TOURNEUR Y., ALLAL L. The symmetry of generalizability theory : Applications to educational measurement.
IRDP - Neuchâtel - 1975 - 33p.

CARDINET J. et ALLAL L. Structure et mesure des objectifs éducatifs dans les surveys périodiques.
IRDP - Neuchâtel - 1976 - 34 p.

CARDINET J., TOURNEUR Y., ALLAL L. The generalizability of surveys of educational outcomes.
IRDP - Neuchâtel - 1975 - 33 p.

CARDINET J. et ALLAL L. Applications de la Théorie de la généralisabilité. Estimations d'erreurs et adaptations de plans de mesure.
Conseil de l'Europe - Strasbourg - 1976

CARDINET Jean Les modèles d'évaluation applicables aux innovations pédagogiques.
IRDP/R 76.19 - Neuchâtel - Octobre 1976

CARDINET J. et ALLAL L. La mesure des résultats de l'enseignement. Compte rendu de l'atelier de contact européen organisé à Windsor du 13 au 18 juin 1976 par le Conseil de l'Europe.
IRDP/R 76.07 - Neuchâtel - Juillet 1976 - 21 p.

CARDINET Jean Pour mieux fonder la mesure en pédagogie.
IRDP/R 76.12 - Neuchâtel - Septembre 1976 - 15 p.

CARDINET Jean L'égalité devant l'examen.
IRDP/R 76.17 - Neuchâtel - 1976 - pp 82 - 93

CARDINET Jean Objectifs éducatifs et évaluation individualisée.
IRDP/R 77.05 - Neuchâtel - Mars 1977 - 57 p.

CRONBACH L. J. Essentials of Psychological testing.
Harper - 1966 - 650 p.

CURETON E.E. Note on φ/φ max.
Psychometrika - 24 - 1959 - pp 88 - 91

DAGNELIE Pierre Théorie et méthodes statistiques (2 Vol.).
Presses agronomiques de Gembloux - 1970 - Gembloux.

DAGNELIE Pierre Analyse statistique à plusieurs variables.
Presses agronomiques de Gembloux - 1975 - Gembloux

DAGNELIE Pierre Le statisticien et l'ordinateur.
Revue Biométrie - Praximétrie - Vol 15 - 1975

DAVIS F. B. Item Analysis Data.
Graduate School of Education - Cambridge (Massachussets) - 1949 - 42 p.

DAWES R. M. Fundamentals of attitude measurement.
J. Wiley & Sons - 1972 - 165 p.

DE CROLY G., THUNUS L., VANDERHAEGEN J., DURIAU J. Evaluation : Les différentes techniques d'élaboration des questions.
Revue de la Direction Générale de l'Organisation des Etudes - 12e année n°3 - Mars 1977 - pp 19 - 28

DE LANDSHEERE G. Comment les maîtres enseignent.
Ministère de l'Education nationale et de la Culture française - Direction générale de l'organisation des études - 1974

DE LANDSHEERE Gilbert Evaluation continue et examens.
Précis de docimologie.
Editions Nathan Labor - Bruxelles - 1972 - 238 p.

DE LANDSHEERE V. et G. Definir les objectifs de l'éducation.
Editions Georges Thone - Liège - 1975 - 293 p.

DE LANDSHEERE G. Pour un lexique des objectifs de l'Education.
Paedagogica Europaea - Vol. XI - 1976 - n°1 - pp 31 - 35

DEMING W. E. The logic of evaluation.
In Handbook of Evaluation Research- STRUENING & GUTTENTAG -
Englewood Cliffs - 1975 - pp 53 - 68

DEPASSE Michel Traitement de données relatives aux effets de deux systèmes éducatifs. Utilisation de l'évaluation informative.
Mémoire de licence non publié - ULB - 1977 - 54 p.

D'HAINAUT Louis Concepts et méthodes de la statistique.
Tome I : Une variable, une dimension.
Editions Labor - Bruxelles - 1975 - 368 p.

DIDAY E. et LEBART L. L'analyse des données.
"La recherche" n°74 - Janvier 1977 - pp 15 - 25

DOWNIE N. M. Fundamentals of Measurement : Techniques and Practices.
Oxford University Press - 1967 - 482 p.

DRESSEL P. L. & Associates Evaluation in Higher education.
Houghton Mifflin C° - Boston - 1961 - 480 p.

EDWARDS A. L. & KILPATRICK F.P. A technique for the construction of attitude scales.
Journal of Applied Psychology - 1948 - Vol 32 - pp 374-384

EISS A. F. Evaluation of instructional systems.
Gordon and Breach - New Yok - 1970 - 139 p.

FAVERGE Jean-Marie L'examen du personnel et l'emploi des tests.
PUF - Le Psychologue - 1972 - 176 p.

FAVERGE Jean-Marie Note sur l'analyse des questions à libellés bipolaires.
Le Travail Humain - Tome 36 n°1/1973 - pp 61 - 74

FAVERGE Jean-Marie Les styles des étudiants.
Psychologica Belgica, 1974, XIV - 2, pp 93 - 104

FAVERGE Jean-Marie Méthodes statistiques en psychologie appliquée. Tome III 2e édition.
PUF - Paris - 1975

FAVERGE Jean-Marie Analyse des images des colonnes d'un tableau de correspondance. Application à un questionnaire de style ouvrier.
Le Travail Humain - Tome 38, n°1/1975 pp 109 - 118

FAVERGE Jean-Marie Correspondance entre l'analyse binaire classique et l'analyse de Benzécri.
Année psychol. - 1977, 77 - pp 149 - 160

FISHBEIN Martin Readings in Attitude, Theory and Measurement.
J. Wiley - 1967 - 499 p.

FRYDMAN Marcel Développement de l'aptitude à exploiter un matériel de références.
Thèse non publiée - Mons - 1975

FRYDMAN Marcel L'utilisation d'un matériel de références.
Recherche expérimentale.
Université de l'Etat à Mons - Département de psychologie - Déc. 1976 - 12 p.

GAGE N. L. Handbook of Research on teaching.
Rand McNally & Cº - Chicago - 1963 - 1218 p.

GAGNÉ R. M. Curriculum research and the promotion of learning.
AERA Monograph series on curriculum evaluation (I)
Rand McNally - New York - 1967 - pp 19 - 38

GLASER R. & NITKO A. Measurement in Learning and Instruction.
In Educational Measurement THORNDIKE L. Second Edition.
American Council on Education - Washington - 1971 - pp 625 - 670

GOOD C. V. Dictionary of Education.
McGraw-Hill - New Yok - 1959 - 679 p.

GOODMAN L.A. & KRUSKAL W. H.
I Measures of association for cross classification.
Journ. Amer. Stat. Assoc. 49 (1954) pp 732 - 764
II Further discussion and references.
Journ. Amer. Stat. Assoc. 54 (1959) pp 123 - 163
III Approximate sampling theory.
Journ. Amer. Stat. Assoc. 58 (1963) pp 310 - 364
IV Simplification of asymptotic variances.
Journ. Amer. Stat. Assoc. 67 (1972) pp 415 - 421

GOSLIN D. A. Teachers and Testing.
Russel Sage Foundation - 1967 - 201p.

GREENE, JORGENSEN & GERBERICH Measurement and Evaluation in the Elementary School.
Longmans, Green & Cº - 1953 - 617 p.

GREEN B. F. Attitude measurement.
In Handbook of Social Psychology (I) - LINDZEY.
Addison-Wesley, Reading - 1954 - pp 335 - 369

GREEN B. F. A method of scalogram analysis using summary statistics.
Psychometrika - Vol. 21, nº 1 March 1956 - pp 79 - 88

GRISAY A. Vers un système d'évaluation formative.
Revue de la Direction générale de l'Organisation des Etudes - nº 5 - Mai 1976 - pp 11 - 28

GRONLUND N. E. Measurement and Evaluation in Teaching.
Mac Millan - New York- 1965 - 420 p.

GUILFORD J. P. The nature of human intelligence.
McGraw-Hill - New York - 1967

GUTTMAN Louis A basis for scaling qualitative data.
American Sociological Review- 1944, 9 - pp 139 - 150

HALLEUX J. et VANDEVELDE L. Quelques variables parasites des dispositifs expérimentaux en pédagogie.
Revue belge de Psychologie et de Pédagogie - Juin 1972 - Tome XXXIV nº138 - pp 45 - 57

HALLEUX J. Contribution à la méthodologie de l'expérimentation en pédagogie. Neutralisation, mise en évidence, contrôle et différenciation de variables parasites.
Thèse non publiée - ULB - 1973

HALLEUX J. Problèmes soulevés par les phénomènes évolutifs en pédagogie expérimentale.
Rev. Sciences de l'éducation - Janv.Mars 1975 - pp 61 - 74

HALLEUX J. Propriétés de l'échantillonnage aléatoire simple.
Rev. Sciences de l'éducation - Janv.Mars 1975 - pp 75 - 83

HALLEUX J. & LEBRUN B. Quelques problèmes de méthodologie générale.
Cours ronéotypé - Service de Didactique expérimentale - ULB - 1975 - 89 p.

HEIDT E. U. Media and Learner Operations : the Problem of a media taxonomy revisited.
British Journal of Educational technology - Vol.8 n°1 - Juin 1977 - pp 11 - 26

HYMAN R. T. Approaches in curriculum.
Prentice-Hall - 1973 - 225 p.

JAMBE R. & MASAI-PERL P. L'éducation de l'affectivité.
Ed. Nathan Labor - Bruxelles - 1976 - 152 p.

KELLEY T. L. The selection of upper and lower groups for the validation of test items.
J. Educ. Psychol. XXX (Janvier 1939) - pp 17 - 24

KING W. H. Statistics in education.
Mac Millan - 1969 - 177 p.

KRATHWOHL D. Taxonomy of Educational Objectives.
Handbook II : Affective Domain.
David McKay - New York, 1964

LEBRUN B. Contraintes didactiques d'une éducation émancipatrice.
Rev. Sciences de l'Education - 1/2 - 1977 - pp 195 - 215

LEROY G. La conquète de l'environnement, l'étude des faits.
Deboeck - Bruxelles - 1975 - 273 p.

LIKERT R. A Technique for the measurement of attitudes.
Archives of Psychology - 1932 - Whole n° 140 - pp 44 - 53

LINDEMAN R. H. Educational measurement.
Scott, Foresman & C° - 1967 - 173 p.

LINDZEY G. & ARONSON E. The handbook of Social Psychology.
Second Edition, Volume two - Research Methods.
Addison - Wesley - Reading - 1968 - 819 p.

LOEVINGER J. The technic of homogeneous tests compared with some aspects of "scale analysis" and factor analysis.
Psychological Bulletin - 1948, Vol. 45 - pp 507 - 529

LUCIER P. Guide pour la mise en route de l'analyse institutionnelle. L'analyse institutionnelle (2).
CADRE - Montréal - 1976 - 46 p.

LYMAN H. B. Test scores and what they mean.
Prentice-Hall - 1963 - 223 p.

MAGER R. F. Comment définir des objectifs pédagogiques.
Gauthier-Villars - Paris - 1974 - 60 p.

MARBAISE C. et DIGHAYE J. A propos d'évaluation.
Revue de la Direction générale de l'Organisation des études - Avril 1975 n°4 - pp 11 - 16

MATALON B. L'analyse hiérarchique.
Gauthier-Villars - Paris - 1965 - 150 p.

MEHRENS W. A. & LEHMAN I. J. Measurement and Evaluation in Education and Psychology.
Holt, Rinehart & Winston - New York - 1973 - 718 p.

MENDES DA COSTA Y. Biologie et didactique expérimentale.
Association nationale des professeurs de biologie de Belgique - n°1, 23e année - mars 1977 - pp 1 - 21

Ministère de l'Education nationale et de la Culture française
Direction générale de l'organisation des Etudes
" Faire le point sur... l'enseignement secondaire rénové ".
Bruxelles - 1974 - 260 p.

Ministère de l'Education nationale et de la Culure française
Circulaire ayant pour objet l'évaluation des études dans l'enseignement secondaire.
2/4/76 - Bruxelles - 14 p.

Ministère de l'Education nationale et de la Culture française
" Enseignement secondaire de type I - aux parents ".
Brochure éditée par la Direction générale de l'Organisation des études - Bruxelles - 1977 (2e édition) - 24 p.

NIE N. & Coll. Statistical package for social sciences (SPSS).
McGraw-Hill - 1975 - 675 p.

NUNNALLY J. C. Tests and Measurement. Assessment and Prediction.
McGraw-Hill - 1959 - 446 p.

OLINER P. M. Teaching Elementary Social Studies.
Harcourt Brace Jovanovich - New York - 1976 - 381 p.

PIDGEON & YATES An Introduction to Educational Measurement.
Routledge & Kegan Paul - 1968 - 117 p.

POPHAM W. J. Criterion-Referenced Measurement.
Educational Technology Publications - 1971 - 108 p.

POPHAM W. J. Evaluating Instruction.
Prentice-Hall - 1973 - 157 p.

RASCH G. An item analysis which takes individual differences into account.
Brit. J. Math. Statist. Psychol. - Vol. 19, 1966

REUCHLIN Maurice Précis de statistique.
PUF - Le Psychologue - 1976 - 256 p.

RIGAUX M. Note concernant la détermination de la valeur discriminative des items.
Revue belge de Psychologie et de Pédagogie - Tome XXIII n° 96 - 1961 - pp 119 - 131

RIGAUX M. Contribution à la solution du problème relatif à la détermination du degré de difficulté des items.
Revue belge de Psychologie et de Pédagogie - Tome XXIV n° 100 - 1962 - pp 105 - 117

RODMELL B. The probable policy effects of monitoring educational performance.
Council of Europe - European Contact Workshop - Windsor, June 1976 - Article ronéotypé - 18 p.

ROSS & STANLEY Measurement in today's schools.
Prentice-Hall - 1954 - 3e édition - 485 p.

SCHOER L. A. L'évaluation des élèves dans la pratique de la classe.
PUF - 1975 - 176 p.

SCRIVEN M. The methodology of evaluation.
AERA Monograph series on curriculum evaluation (I)
Rand McNally - New York - 1967 - pp 39 - 83

SKINNER B. F. La révolution scientifique de l'enseignement.
Dessart - Bruxelles - 1968 - 314 p.

SOKAL R. R. & SNEATH P.H.A. Principles of numerical taxonomy.
Freeman - San Francisco - 1963 - 359 p.

SPADA H. Measurement as an Element of Curricular evaluation.
Studies in Educational Evaluation - Vol. 2 n°2 1976
pp 85 - 102

STAKE R.E. Toward a technology for the evaluation of educational programs.
AERA Monograph series on curriculum evaluation (I)
Rand McNally - New York - 1967 - pp 1 - 12

STODOLA Q. & STORDAHL K. Basic educational Tests and Measurement.
SRA - 1967 - 310 p.

STUFFLEBEAM et Coll. Educational Evaluation and Decision Making.
Peacock - Itasca (Illinois) - 1971 - 368 p.

THOMAS R. M. Judging student progress.
David McKay C° - New York - 1954 - 518 p.

THORNDIKE & HAGEN Measurement and Evaluation in Psychology and Education.
John Wiley & Sons - 1955 - 575 p.

TOURNEUR Y. et CARDINET J. La recherche sur les problèmes du contrôle des normes nationales de réussite scolaire.
Conseil de l'Europe - Strasbourg - 1976

TOURNEUR Y. et FRYDMAN M. Une application de la théorie de la généralisabilité à l'estimation d'erreurs.
Université de l'Etat à Mons - Doc 760.620 - 1977 - 17 p.

TRAVERS R.M.W. Second handbook of research in teaching.
Rand McNally - 1973 - 1400 p.

TYLER R. W. Changing Concepts of Educational Evaluation.
AERA Monograph series on curriculum evaluation (I)
Rand McNally - New York - 1967 - pp 13 - 18

VANNESSAG Nadine Démystification de l'autoévaluation.
Mémoire non publié - ULB - 1977

VANDER ELST Pierre Prolégomènes à une pédagogie possible.
Thèse non publiée - ULB - 1976 - 281 p.

VANDEVELDE L. Etude critique des dispositifs expérimentaux en pédagogie.
Revue des Sciences de l'Education - 2/3 1973 - pp 59 - 72

VANDEVELDE L. & VANDER ELST P. Peut-on préciser les objectifs en éducation ?
Nathan Labor - Bruxelles - 1975 - 208 p.

VANDEVELDE L. & LEBRUN B. Essai d'évaluation scolaire informative.
Revue belge de Psychologie et de Pédagogie - n°152
Déc. 1975 - pp 77 - 98

VANDEVELDE L. et GUISLAIN G. Lexique relatif au cours de pédagogie psychologique et expérimentale.
Presses Universitaires de l'ULB - 2e édition - 1977 - 39 p.

WEISS C. H. Evaluation Research in the Political Context.
In Handbook of Evaluation Research - Struening & Guttentag
Sage Publications - Beverly Hills - Vol. I - 1975 -
pp 13 - 25

WHEELER D. K. Curriculum Process.
University of London Press - London - 1967 - 320 p.

WITTROCK M. C. & WILEY D. E. The Evaluation of Instruction.
Issues and Problems.
Holt, Rinehart and Winston - 1970 - 494 p.

WOOD R. & SKURNIK L. S. Item Banking.
NFER - Slough - 1969 - 164 p.

YOUDEN W. J. Index for rating diagnostic tests.
Cancer - A journal of the American Cancer Society -
Vol. III - 1956 - pp 32 - 35

ANNEXE 1

****

TABLE DES MATIERES

RECHERCHE SYSTEMATIQUE DE LONGUE DUREE EN COLLABORATION AVEC DES ENSEIGNANTS DE MATHEMATIQUE

	Pages
INTRODUCTION	1
I TRAVAUX PRELIMINAIRES	2
1) Année scolaire 1971-1972	2
2) Année scolaire 1972-1973	4
3) Année scolaire 1973-1974	8
II BASES DU TRAVAIL EN COLLABORATION	12
III APPORTS DU SERVICE DE DIDACTIQUE EXPERIMENTALE	14
1) Année scolaire 1974-1975	14
2) Année scolaire 1975-1976	14
3) Année scolaire 1976-1977	16
IV TECHNIQUES D'ANALYSE EMPLOYEES	17
1. Présentation des données sur cartes perforées	17
2. Programmes d'analyse selon la technique proposée au professeur (METHRØ)	19
a) limites du programme de base	19
b) vérification des données	19
c) informations figurant sur le listing	19
d) exemple de listing	21
e) programmes dérivés	22
3. Programmes d'analyse d'une population plus vaste (METSCI)	22
a) limites du programme de base	22
b) vérification des données	22
c) informations figurant sur le listing	22
d) programmes dérivés	23

Pages

4. Programmes d'analyse comparative de la population répartie en classes 23

A. Etude de l'homogénéité de la population pour chaque item (arc sinus) 23

1. Programme de transformation des données (PREPSI9) 24

2. Programme d'analyse par la technique "arc sinus" (ARCSIN9) 24

a) limites du programme 24

b) algorithme employé par l'ordinateur 24

c) informations figurant sur le listing 25

d) exemple de listing 26

B. Etude de l'homogénéité pour un ensemble d'items (XHI^2) 27

1) Utilisation du programme sans écarter aucune classe ni aucun item 27

2) Utilisation du programme en écartant des classes et/ou des items 27

3) Exemple de listing 28

5. Programmes d'évaluation hiérarchique 28

A. Programme de calcul des matrices de concordance (CØIDSO9) 28

1) Limites du programme 28

2) Informations figurant sur le listing 28

3) Exemples de listings 30

B. Programme de hiérarchisation des items (ITEMSF9) 31

1) Limites du programme 31

2) Informations figurant sur le listing 31

6. Programmes d'analyse des données 32

a) Transformation des données en vue de l'analyse typologique (PREPTY9) 32

b) Analyse des correspondances par groupe typologique (ABINOO9) 33

c) Analyse des correspondances par individu (ABINOO9) 33

d) Transformation des données par classe en vue de l'analyse des correspondances (PREBZO9) 33

Pages

V EVOLUTION DES EPREUVES ET ANALYSE DES RESULTATS 34

a) Evaluation informative des résultats de chaque classe 34

b) Evaluation informative des résultats du groupe de classes 35

VI ETUDE DE L'EPREUVE N°1 36

A. Projet de 1973-1974 36

I. Mémoire 37

II. Utilisation 38

III. Compréhension 39

Cotation du test n°1 41

Analyse des résultats 42

1°) Evaluation pour l'ensemble du groupe 42

2°) Homogénéité des classes par rapport à l'ensemble de l'épreuve (XHI^2) 43

3°) Stabilité des items (arc sinus) 44

4°) Analyse des correspondances 45

a) Analyse sur base typologique 45

b) Analyse par classe 47

B. Contrôle de 1974-1975 (M41) 48

I Mémoire 49

II Compréhension 50

III Utilisation 52

Barême de correction du contrôle 54

Analyse des résultats 55

1°) Evaluation par classe 55

2°) Evaluation pour l'ensemble du groupe 55

3°) Homogénéité des classes par rapport à l'ensemble de l'épreuve (XHI^2) 57

4°) Homogénéité des classes par rapport à chacun des subtests (XHI^2) 59

5°) Homogénéité des classes par rapport à chaque item (arc sinus) 60

6°) Recherche d'une hiérarchie 61

a) Etude de la concordance des items 61

b) Analyse hiérarchique 61

Pages

7°) Analyse des correspondances 63
a) sur base typologique 63
b) par classe 63
C. Contrôle de 1975-1976 (M61) 66
1°) Evaluation par classe 66
2°) Evaluation pour l'ensemble du groupe 66
3°) Homogénéité des classes par rapport à un groupe d'items (XHI^2) 67
a) par rapport à l'ensemble de l'épreuve 67
b) par rapport à chaque subtest 68
4°) Homogénéité des classes par rapport à chaque item (arc sinus) 68
5°) Recherche d'une hiérarchie 69
a) Etude de la concordance des items 69
b) Analyses hiérarchiques 69
6°) Analyse des correspondances 70
a) sur base typologique 70
b) sur les 15 classes 70.
D. Contrôle de 1976-1977 (M81) 73
I Mémoire 74
II Compréhension 75
III Utilisation 77
Analyse des résultats 79
1°) Evaluation par classe 79
2°) Evaluation pour l'ensemble du groupe 79
3°) Homogénéité des classes par rapport à l'ensemble de l'épreuve (XHI^2) 80
4°) Stabilité des items (arc sinus) 81
5°) Recherche d'une hiérarchie 82
a) Etude de la concordance des items 82
b) Analyse hiérarchique 83
6°) Analyse des correspondances 85
a) sur base typologique 85
b) sur les 17 classes 85

Pages

7°) Etude systématique par subtests 88
a) Etude par classe pour chaque subtest 88
b) Etude pour l'ensemble de la population pour chaque subtest 89
c) Homogénéité des classes par rapport à chacun des subtests (XHI^2) 90
d) Analyse des correspondances 90
1) sur base typologique 90
2) sur les 17 classes 95
E. Evolution au fil des années 99
VII ETUDE DE L'EPREUVE N°2 100
A. Travaux préparatoires 100
1) Projet initial 100
2) Modifications intervenues 106
B. Forme utilisée en 1975-1976 107
Analyse des résultats 113
1°) Evaluation au niveau de chaque classe 113
2°) Evaluation pour l'ensemble de la population 114
3°) Homogénéité des classes par rapport à l'ensemble de l'épreuve (XHI^2) 115
4°) Stabilité des items (arc sinus) 116
5°) Recherche d'une hiérarchie 117
a) Etude de la concordance des items 117
b) Analyse hiérarchique 118
6°) Analyse des correspondances 119
a) Analyse sur base typologique 119
b) Analyse sur les individus 123
c) Analyse des résultats par classe 125
C. Epreuve proposée en 1976-1977 127
Analyse des résultats 134
1) Evaluation au niveau de chaque classe 134
2) Evaluation de l'ensemble de la population 134

Pages

3) Homogénéité des classes par rapport à l'ensemble de l'épreuve (XHI^2) 135
4) Stabilité des items (arc sinus) 136
5) Recherche d'une hiérarchie 137
a) Etude de la concordance des items 137
b) Analyse hiérarchique 137
6) Analyse des correspondances 139
a) Analyse sur base typologique 139
b) Analyse sur les individus 140
c) Analyse par classe 140

VIII ETUDE COMPARATIVE DES RESULTATS DE TOUTE UNE ANNEE SCOLAIRE 144

Présentation de l'étude 144
Tableau comparatif 145
Commentaires 146
Conclusions 148

CONCLUSIONS GENERALES 149

TABLE DES MATIERES 150

ANNEXE 2

****

TABLE DES MATIERES

APPLICATION DES TECHNIQUES D'EVALUATION INFORMATIVE A LA CONSTRUCTION D'EPREUVES STANDARDISEES

	Pages
INTRODUCTION	1
I PRESENTATION DES EPREUVES	2
1°) Epreuve pronostique de compréhension	3
1. Objectif	3
2. Contenu	3
3. Dispositif	3
4. Résultats	3
2°) Epreuve diagnostique de connaissance	3
1. Objectif	3
2. Contenu	3
3. Dispositif	4
4. Résultats	5
3°) Epreuve pronostique d'application	5
II ETUDE DES PREMIERS RESULTATS OBTENUS	7
A. Epreuve pronostique de compréhension : Les distances	7
- Projet initial (1975 - 1976)	7
a) Evaluation informative sur l'ensemble des élèves	14
b) Analyse des correspondances sur base typologique	14
Conclusions	16
- Epreuve proposée en juin 1977	17
B. Epreuve diagnostique de connaissance	24
Test 1	25
Test 2	28
Test 3	31
Test 4	36
Analyse des résultats enregistrés	42

Pages

1. Test 1 43.
2. Test 2 44
3. Test 3 45
4. Test 4 46

Conclusions 47

III CARACTERISTIQUES DES EPREUVES PROJETEES 47

1°) L'élaboration d'une épreuve diagnostique pour le début de l'année 47

2°) L'introduction d'un nombre restreint de ces questions dans une épreuve d'utilisation de l'acquis 48

IV CONTENU DES EPREUVES PROJETEES

A. Epreuve diagnostique 49
- Caractéristiques des questions retenues 51

B. Epreuve d'utilisation de l'acquis 54
- Caractéristiques des questions retenues 55

CONCLUSIONS 56

TABLE DES MATIERES 57

ANNEXE 3

****

TABLE DES MATIERES

PRESENTATION SIMPLIFIEE DE L'EVALUATION INFORMATIVE

DANS LE CADRE DE LA FORMATION DES ENSEIGNANTS

	Pages
INTRODUCTION	1
I SCHEMA D'ANALYSE DE L'EVALUATION INFORMATIVE	3
1. Qu'est-ce qu'une matrice de résultats ?	3
2. Exemple de matrice de résultats	3
3. Evaluation des résultats des élèves	4
a) Calcul de la moyenne	4
b) Calcul de l'écart-type	5
c) Calcul des limites sigmatiques	5
Histogramme de répartition des élèves	6
4. Evaluation du comportement des items	7
a) Les items supérieurs	7
b) Les items inférieurs	7
c) Les items centraux	9
d) Les items non concordants	10
e) Un cas particulier : le contraste	12
f) Conclusion	13
1. L'observation confirme les ambitions	13
2. L'observation infirme les ambitions	13
3. L'observation n'est pas de nature à caractériser le résultat	14
II TECHNIQUE D'EVALUATION INFORMATIVE	15
1. La loi binomiale des probabilités	15
Table d'intervalle autour de 50 % de réussite	17
2. Utilisation de la table	17
3. Première conclusion	18

Pages

4. Recherche d'une différenciation 19
a) Calcul de la note "provisoire" (NP) 20
b) Calcul de la note de l'élève "moyen" ($\bar{x}e$) 20
c) Calcul du pouvoir de différenciation de chaque item 21
Exemple 23
d) Calcul de la note de différenciation (ND) 24
5. L'interprétation des résultats 24
6. Tableau schématique de la technique 25

III CONCLUSIONS ET SYNTHESE 26

IV EXEMPLE D'APPLICATION DE LA TECHNIQUE D'EVALUATION INFORMATIVE 27

a) Répartition des élèves en trois catégogies 29
b) Calcul de la note provisoire (NP) par élève 30
c) Calcul de la note de l'élève moyen 31
d) Répartition des élèves par rapport à la moyenne 31
e) Calcul de l'IDS 32
f) Calcul de la note de différenciation (ND) 33
Matrice analysée sans ordonner les élèves et les items 34
Distribution des notes ordinales sur l'ensemble des items 35
Distribution des notes donnant lieu à différenciation 35
Listing des résultats enregistrés 36
Exercice 38
Correction de l'exercice 39
Interprétation 40

Pages

V APPLICATION DE LA TECHNIQUE D'EVALUATION INFORMATIVE

INTERPRETATION DE RESULTATS 41

Exercice n°1 : Les Gorges de l'Ardêche 41

Source 41

Test 42

Objectifs visés par l'épreuve 46

Population 47

Circonstances 47

Matrice des résultats 48

Listing des résultats analysés 49

Interprétation 50

1) Items d'enquête supérieure 50

2) Items de différenciation 50

3) Items d'enquête inférieure 50

Conclusion 50

Exercice n°2 : XVZ 51

Information 52

Exercices 53

Matrice des résultats (1ère latine) 54

Matrice des résultats (3ème latine) 55

TABLE DES MATIERES 56

ANNEXE 4

****

TABLE DES MATIERES

ETUDE COMPARATIVE DE DONNEES EXPERIMENTALES SELON DIFFERENTES METHODES D'ANALYSE

	Pages
INTRODUCTION	1
I SITUATION DE L'EXPERIENCE	2
Test " Les Gorges de l'Ardêche "	3
II ANALYSE DES RESULTATS	7
A. Evaluation informative	7
1. Etude de l'ensemble de la population	7
Interprétation	7
a) Items d'enquête supérieure	7
b) Items de différenciation	7
c) Items d'interaction	8
Conclusion	9
2. Evaluation hiérarchique	10
a) Matrice d'interconcordance	10
b) Etude des hiérarchies	11
1. Etude des couples très concordants	12
2. Etude du groupe de 3 items interconcordants	12
3. Etude du groupe de différenciation	12
B. Application des méthodes d'analyse concurrentes	13
1. Analyse des correspondances	13
Répartition typologique	13
Analyse des correspondances (10 groupes)	14
Analyse des correspondances (8 groupes)	15
2. Analyse binaire	16
Représentation brute	18
Interprétation des résultats de l'analyse binaire	19
CONCLUSIONS	20
TABLE DES MATIERES	21

ANNEXE 5

****

TABLE DES MATIERES

UTILISATION DES METHODES D'EVALUATION INFORMATIVE DANS DES RECHERCHES EXPERIMENTALES

	Pages
INTRODUCTION	1
1° RECHERCHES DE BAMBA	2
A. Etude de l'homogénéité du groupe expérimental	3
1. Technique XHI^2	3
a) Vérification de l'homogénéité de l'ensemble de la population	3
b) Etude de l'homogénéité des sous-groupes	4
1. Sous-groupe lingalaphone	4
2. Sous-groupe francophone	4
2. Technique arc sinus	5
3. Technique XHI^2 classique	5
B. Etude des épreuves expérimentales	6
1. Contenu des épreuves	6
a) épreuves de vocabulaire (1 et 2)	6
b) épreuve de lecture (3)	6
c) épreuve de calcul (4)	7
d) épreuve de langage oral (5, 6, 7)	7
e) épreuve de compréhension verbale (8)	7
2. Analyse des résultats	7
a) épreuves de vocabulaire	7
- Liste des items du test 1	7
- Liste des items du test 2	8
- Comportement de l'ensemble de la population et des différents sous-groupes	9
- Comportement des classes séparées	10
- Etude des résultats par la XHI^2 classique	11

Pages

b) épreuve de lecture 11
- Liste des items du test 3 11
- Comportement de l'ensemble de la population et des différents sous-groupes 12
- Comportement des classes séparées 12
- Etude des résultats par la XHI^2 classique 13

c) épreuve de calcul 13
- Liste des items du test 4 13
- Comportement de l'ensemble de la population et des différents sous-groupes 13
- Comportement des classes séparées 14
- Etude des résultats par la XHI^2 classique 14

d) épreuve de langage oral 15
- Liste des items du test 5 (production verbale) 15
- Comportement de l'ensemble de la population et des différents sous-groupes 15
- Comportement des classes séparées 16
- Etude des résultats par la XHI^2 classique 17

- Présentation des items du test 6 (correction syntaxique) 17
- Comportement de l'ensemble de la population et des différents sous-groupes 17
- Comportement des classes séparées 17
- Etude des résultats par la XHI^2 classique 18

- Présentation des items du test 7 (maniement oral de la langue) 18
- Comportement de l'ensemble de la population et des différents sous-groupes 18
- Comportement des classes séparées 19

Pages

e) épreuve de compréhension verbale 19
- Liste des items du test 8 20
- Comportement de l'ensemble de la population et des différents sous-groupes 20
- Comportement des classes séparées 22
- Etude des résultats par la XHI^2 classique 23

2° RECHERCHES DE FRYDMAN 24

1) Objectifs de la recherche 24
2) Dispositif expérimental 24
- Représentation schématique du dispositif 25
3) Traitement 26
4) Posttests 27
5) Résultats 27

A. Constitution des échantillons expérimentaux 28
1. Analyse du groupe total 28
- Tableau synthétique des observations sur l'ensemble de la population 29
2. Analyse des sous-groupes 30
a) Etude de la cohérence interne des sous-groupes 30
- Groupe A 30
- Groupe B 30
- Groupe C 31
- Conclusions 31
b) Applications d'autres techniques 31

B. Analyse des résultats du traitement 32
1) Première épreuve 33
2) Deuxième épreuve 34
3) Troisième épreuve 35
4) Quatrième épreuve 36
5) Cinquième épreuve 37
6) Sixième épreuve 38

Pages

7) Evaluation hiérarchique 39
a) Matrices de concordance 39
Groupe A 39
Groupe B 40
b) Etude des hiérarchies 41
8) Application d'autres techniques 42
9) Conclusions 42
C. Analyse des résultats aux posttests 43
1. Homogénéité du groupe total 43
- Interprétation 44
- Comparaison pour chacun des subtests 44
1er subtest 44
2e subtest 45
3e subtest 45
2. Homogénéité du groupe A 46
3. Homogénéité du groupe B 47
4. Homogénéité du groupe C 48
5. Homogénéité du groupe D 49
6. Evaluation hiérarchique 50
a) Matrice de concordance 50
b) Etude des hiérarchies 50

3° RECHERCHES D'AMEGAN 51

A. Objectifs poursuivis par Samuel Amégan 51
B. Tâches choisies et schéma expérimental 52
1. Echantillon 52
2. Tâches 52
1) Le prétest 52
a) Torrance figuratif forme A 53
b) Torrance verbal forme A 53
2) Le traitement 54

Pages

3) Le posttest 54
a) Torrance figuratif forme A 54
b) Torrance verbal forme A 54
c) Guilford production divergente 54
3. Tableau résumé du schéma expérimental 54
C. Objet du mémoire de M. DEPASSE par rapport à la recherche de Samuel Amégan 56
D. Résultats observés par M. DEPASSE 57
1. Dichotomisation des données 57
a) Principes retenus 57
b) Choix du seuil 57
c) Conséquences de ce choix 57
2. Evaluation informative 58
a) Etude des résultats globaux 58
Conclusion 58
b) Comparaison des classes "actives" et "conventionnelles" 59
Résultats au Torrance figuratif 59
Résultats au Torrance verbal 59
Conclusions 59
c) Etude des résultats au Guilford production divergente 60
Conclusion 60
d) Attribution des effets constatés 60
Torrance figuratif 61
Torrance verbal 61
Guilford production divergente 61
Conclusion 61
e) Comportement des items aux tests de Torrance 62
1. Interprétations proposées 62
2. Tableau récapitulatif des observations 62
3. Conclusions 63

CONCLUSIONS 64

TABLE DES MATIERES 65

TABLE GENERALE DES MATIERES

****

UNE EVALUATION INFORMATIVE

Essai de mise au point de méthodes
susceptibles de rendre compte
d'effets pédagogiques collectifs

Pages

EXPOSE DE LA THESE - HYPOTHESES FONDAMENTALES 1

INTRODUCTION : Situation de l'évaluation
dans l'action pédagogique 4

Mise au point terminologique 4
Schéma général de l'action éducative 4
1. LA DEFINITION DES OBJECTIFS 4
2. LA MISE EN OEUVRE D'ACTIVITES DIDACTIQUES 6
3. L'EVALUATION 9
1°) La prise de mesure 11
2°) L'analyse des données 13
3°) La décision 16
4°) La communication 18
Résumé schématique de l'action éducative 20

CHAPITRE I : FONDEMENTS THEORIQUES DE NOTRE PROPOSITION
D'EVALUATION INFORMATIVE 21

1°) La typologie psychométrique des tests 21
1. Le test de classement 22
2. Le test d'enquête 23
3. Le test d'inclination 24
4. Le test hiérarchique 25
2°) Le scallogramme 26

Pages

CHAPITRE II : PRINCIPES DE BASE DE L'EVALUATION INFORMATIVE 28

1. Contraintes imposées par nos hypothèses fondamentales 28
   a) Choix d'une méthode valide 28
   b) Choix d'une méthode commode 29
   c) Conclusion 30
2. Caractéristiques générales de nos méthodes d'analyse 30
3. Choix d'un mode de notation 33
4. Choix d'un critère de concordance 35

CHAPITRE III : SCHEMAS D'ANALYSE DE L'EVALUATION INFORMATIVE 38

I La correspondance entre un niveau atteint et une ambition annoncée 39

   1°) Présentation du problème : la catégorisation des items 39
      1. L'observation confirme les ambitions 39
      2. L'observation infirme les ambitions 39
      3. L'observation n'est pas de nature à caractériser le résultat 39
   a) Exploitation du pourcentage de réussite 40
   b) L'opposition entre deux sous-groupes d'élèves 42
   2°) Moyens techniques proposés 44
   A. Effectifs compris entre 10 et 30 élèves 44
      1. Présentation de la technique 44
      - TABLE D'INTERVALLE AUTOUR DE 50 % DE REUSSITE 44
      - Algorithme pour la détection des items de différenciation 45
      - L'Indice de Différenciation Spécifique (IDS) 46
      - L'interprétation 47

Pages.

2. Commentaires sur la technique proposée 47
a) Recours à la loi binomiale des probabilités pour construire la table d'intervalle 47
b) Sommation des notes enregistrées à 50 % de réussite 48
c) Dichotomisation des notes provisoires 49
d) L'indice de différenciation spécifique 50
3. Tableau schématique de la technique 51
B. Extension à des populations plus vastes 52
1) Présentation de la technique 52
a) L'intervalle de confiance autour de 50 % de réussite 52
- Evolution de la binomiale 53
b) Le critère d'acceptation des items de différenciation 56
2) Tableau schématique de la technique 58
C. Conclusion 59

II La mise en évidence de hiérarchies 60

1°) Présentation du problème : le classement des activités 60
a) Les techniques destinées à graduer les attitudes 63
b) Les taxinomies utilisées dans les sciences de la nature 65
c) L'analyse binaire 66
d) La typologie psychométrique des tests 67
e) Conclusions 68
2°) Les moyens techniques proposés 69
A. Recherche de la base d'une hiérarchie : La famille-critère 69
1. L'indice de similarité 69
2. La similarité maximale 71
3. La concordance 71

Pages.

4. La matrice de similarité 73
5. La matrice symbolique de concordance 73
6. La sélection de "familles-critères" 73
B. La hiérarchisation des items 74
1. Dichotomisation de la "famille-critère" 74
2. Calcul de la concordance 74
3. Sélection des items hiérarchisés 74
4. Interprétation 74
3°) Commentaires sur la technique proposée 75
a) Caractéristiques de l'épreuve hiérarchique 75
b) L'indice de concordance 76
1. Coefficient d'homogénéité entre deux items 78
2. Coefficient d'homogénéité d'un test 79
c) La hiérarchie autour d'un critère 80
d) Les étapes de la recherche d'une hiérarchie 82
4°) Tableaux schématiques de la technique 83
a) Recherche de "familles-critères" 83
b) Hiérarchisation des items 84
5°) Conclusions 84
III Influence des institutions scolaires sur les effets collectifs 85
1°) Présentation du problème : analyse d'une population répartie en classes 85
2°) Moyens techniques proposés 86
1. Stabilité de la population pour un item isolé (technique "ARC SINUS") 86
a) Transformation de la proportion de réussite totale 87
b) Limites angulaires de l'intervalle de confiance 87
c) Proportions marquant les limites de l'intervalle de confiance 87

Pages.

2. Stabilité de la population pour un groupe d'items (technique "XHI^2") 89
3°) Commentaires sur les techniques proposées 90
a) La technique "ARC SINUS" 90
1. L'effet de contexte 91
2. L'apport de chaque individu 91
b) La technique "XHI^2" 91
4°) Tableaux schématiques des techniques 92
a) Technique "ARC SINUS" 92
b) Technique "XHI^2" 93
5°) Conclusions 94

CHAPITRE IV : MISE A L'EPREUVE DES METHODES D'EVALUATION INFORMATIVE 95

Annexe 1 Recherche systématique de longue durée en collaboration avec des enseignants de mathématique 95
Annexe 2 Application des techniques d'évaluation informative à la construction d'épreuves standardisées 96
Annexe 3 Présentation simplifiée de l'évaluation informative dans le cadre de la formation des enseignants 96
Annexe 4 Etude comparative de données expérimentales selon différentes méthodes d'analyse 97
Annexe 5 Utilisation des méthodes d'évaluation informative dans des recherches expérimentales 97

CONCLUSIONS 98

Pages.

BIBLIOGRAPHIE DE TRAVAIL 104

TABLES DES MATIERES 120

Annexe 1 120
Annexe 2 126
Annexe 3 128
Annexe 4 131
Annexe 5 132

TABLE GENERALE DES MATIERES 137