L'EVALUATION EN PEDAGOGIE

Du bon usage des tests

Daniel PASQUIER

EAP

– *Chez le même éditeur* –

Collection : Sciences Humaines

animée par T. EGO - 1E -

R. MUCCHIELLI - Analyse et liberté (1E01)
E. GUILMAIN, G. GUILMAIN - L'activité psychomotrice de l'enfant (1E02)
J.C. WITTERSHEIM - La prévention des accidents par les méthodes psychologiques (1E03)
R. BONNARDEL - Intelligence concrète et psychométrie (1E04)
S. LIEVENS, M. CLIQUET - La sélection du personnel aujourd'hui (1E05)
J. THOMAS - La gestion du personnel aujourd'hui (1E06)
C. CHANDESSAIS - Introduction à l'étude des systèmes (1E07)
G. VERRIEST, G. HERMANS - Vue et Profession, les aptitudes visuelles professionnelles (1E08)
P. GOGUELIN - Psychologie du travail: perspectives 1990 (1E90)
P. GOGUELIN - Psychologie du travail et société post - industrielle (1E91)
P. GOGUELIN - Psychologie du travail : nouveaux enjeux (1E92)

Collection : Tests Psychologiques

animée par R. SIMONNET - 1K -

R. MUCCHIELLI - Le test du Village Imaginaire
A. ABRAHAM - Le dessin d'une personne
A. ABRAHAM - Le monde intérieur des enseignants
J. ROYER - Le test des Contes
J. ROYER - Le dessin d'une maison
P. LAURENT, P. HUSSON - L'insertion du débile mental dans le monde du travail
C. BIEDMA, P. D'ALFONSO - Le langage du dessin (test de Wartegg - Biedma)
G. WEINSTEIN - L'auto - évaluation et les stades de l'ombre
A. REY - Interprétation de dessins et développement psychologique
A. REY - Examen clinique en psychologie et techniques psychométriques
G. BOULANGER BALLEYGUIER - La formation du caractère pendant les premières années
O. BRUNET, I. LEZINE - Le développement psychologique de la première enfance
R. ZAZZO - La nouvelle échelle métrique de l'intelligence
J. COTTRAUX - Méthodes d'évaluation des comportements
J.J. GUILLARME - Le test "Histoire d'une gomme"
J.J. GUILLARME - Evaluation des aides à l'enfant

Collection : Recherches et Sciences de l'éducation

animée par L. MARMOZ - 1G01 -

L. MARMOZ - Les sciences de l'éducation en France

L'EVALUATION

EN PEDAGOGIE

Du bon usage des tests

Du même auteur

D. PASQUIER :

TISS – Test d'intérêt pour les situations scolaires

TPCP – Test prédictif pour le cours préparatoire

TLCP – Test de lecture pour le cours préparatoire

TMCE – Test de mathématiques pour le cours élémentaire

TP6 – Test pour le passage en sixième

TECM – Test d'évaluation des capacités métacognitives

CO – Comprendre l'échec scolaire, essai théorique de psychologie de l'éducation

D. PASQUIER, J. THEBAULT :

TMCP – Test de mathématiques pour le cours préparatoire

D. PASQUIER, P. CHAMAULTE :

TFCE – Test de français pour le cours élémentaire

EDITIONS E.A.P.
6 bis, Rue André Chénier
92130 ISSY - LES - MOULINEAUX

L'EVALUATION EN PEDAGOGIE

- Du bon usage des tests -

par

Daniel PASQUIER

Docteur en sciences de l'éducation

EDITIONS EAP

6 bis, rue A. CHENIER
92130 – ISSY-LES-MOULINEAUX
FRANCE

2ème trimestre 1990
Dépôt légal : 315 – 0390 – 1D09
ISBN : 2.86491 – 074 – 8

TABLE DES MATIERES

SUMMARY

Evaluation is in a crisis. Everybody knows it, everybody tells it, everybody writes it, everybody deplores it. Of course no miracle cure of this central problem of initial education and continuous formation exists. It is not a sufficient reason to do nothing.

Into a first part, the author tries to draw the outlines of what evaluation is, analyses its functions and discloses its mechanisms. The different frameworks of evaluation critique are reviewed and elaborated upon. The yet utopian alternative of formative evaluation is described in detail. A first way to initiate a positive change, a significant technical improvement, a decentration beyond routines is suggested as reasoned and rational use of pedagometric tests.

Therefore, the book's target is to suggest a theoritical thought that catches the assessment problematic bases and to extend this thought towards the practical level by a moderated and justified use of scientific assessment techniques.

Mixing thought and concrete information will make it easier for school or vocational assessment and pedagogy practicians who feel the need for a change, differenciation, widening, enrichment and improvement of assessment practices.

TABLE OF CONTENTS

AVERTISSEMENT

Ce livre a été composé pour compléter l'information des professionnels de la pédagogie, les instituteurs, les professeurs, les rééducateurs, les orthophonistes, les animateurs, les moniteurs, les formateurs, les responsables de formation, les conseillers des Centres de bilan interinstitutionnel, les coordinateurs de zone du Crédit formation, les conseillers des Missions locales et des Permanences d'accueil et d'orientation, les directeurs d'établissements... sur la problématique de l'évaluation en général et sur les possibilités d'application offertes par l'utilisation des tests pédagométriques en particulier. Il contient tous les éléments utiles au démarrage d'une pratique évaluative qui inclut l'usage des tests dans sa démarche.

INTRODUCTION

La crise de l'enseignement est un leit-motiv des journalistes de l'éducation depuis au moins trente ans. Parmi les nombreux sous-ensembles qui composent les ingrédients de cette crise pour laquelle on ne perçoit toujours pas "le bout du tunnel" on peut citer en vrac et de manière non exhaustive: le réformisme ministériel de façade, l'imprévision des flux démographiques, l'échec scolaire, les rythmes scolaires aberrants, le surmenage des élèves et des professeurs, le manque de crédits, l'illettrisme, la démocratisation du collège et du lycée mal conduites, les filières élitiques camouflées, la négligence de la formation technique, le non-renouvellement des méthodes, l'absence de formation psychopédagogique des enseignants, la pauvreté des moyens pédagogiques, l'inadaptation des formations aux besoins immédiats des entreprises, l'incapacité de vivre avec son temps, l'incapacité de se moderniser et de se professionnaliser, le manque d'une politique ambitieuse de recherche fondamentale en pédagogie, le manque de considérations sociale et économique du métier d'enseignant, l'absence de plans de carrière stimulants, la chute des "vocations pédagogiques"... et aussi **la crise de l'évaluation.**

On peut situer la crise de l'évaluation sur deux plans principaux. D'une manière plutôt philosophique voire d'un point de vue politico-idéologique, on a pu constater à partir de mai 68 une tendance à refuser de juger l'autre qui a pu se prolonger concrètement dans des formes d'idéalismes pédagogiques généreux souvent hâtivement assimilés à du laxisme, mais qui n'ont pas su s'armer des méthodes et des outils pédagogiques pertinents. Même les autorités ministérielles ont encouragé ce mouvement en supprimant la notation chiffrée et en la remplacant par une échelle alphabétique. A l'inverse de ces attitudes en retrait par rapport à la problématique évaluative, attitudes dont les promoteurs s'illusionaient en pensant qu'un problème de cette nature se règlait par la suppression pure et simple des pratiques, d'autres idéologues, confondant la rigidité et la rigueur, défendent la conception d'une pratique de l'évaluation sans faiblesse ni pitié. Il en résulte des effets de sursélection et à titre d'exemple on peut citer la pénurie en France de mathématiciens et de scientifiques. Sur un plan psychologique les conséquences de ces attitudes extrêmes sont identiques.

La première se manifeste dans toutes les formes de déresponsabilisation observées ici et là chez les élèves: à quoi bon travailler alors que toutes mes productions sont mises au même niveau que celles des autres, qu'elles ne sont pas estimées à leur juste valeur, qu'elles ne sont pas prises en considération...; à quoi bon travailler lorsque je suis certain de récupérer une note dérisoire, lorsque mes efforts ne seront pas reconnus, lorsqu'on me signifiera une fois de plus mon absence de don... Dans les deux cas, la non prise en compte du travail fourni, les retours négatifs ou indifférents mais surtout indifférenciés risquent d'induire les désinvestissements scolaires et formatifs.

La seconde conséquence agit au niveau de la régulation fonctionnelle des apprentissages, en tant que perturbation de cette régulation. On sait bien, depuis l'invention de la cybernétique et le développement du concept de système, que toute activité se régule par la prise en compte de ses effets. Lorsqu'il n'y a pas d'effet en retour ou bien lorsque cet effet est toujours identique à lui-même, l'activité risque de disparaître, de tourner sur elle-même comme un rituel dont les symboles auraient perdu toute signification, de dériver puis de se perdre car sans but. La répétition des effets négatifs, quant à elle, provoque le désintérêt et qui plus est les réactions de fuite et d'évitement. Dans les deux cas, l'empêchement d'une régulation des apprentissages saine, justifiée, aux objectifs explicités et partagés, et efficace car centrée sur l'apprenant, cet empêchement débouche d'une manière irrémédiable sur la désorientation et la désorganisation psychologiques des élèves et des stagiaires.

La troisième conséquence peut s'exprimer en termes de déscolarisation. L'évaluation indifférenciée par son manque de rigueur ou par son excès participe, à côté de nombreux autres facteurs, à la perte de confiance des élèves et des stagiaires dans les appareils de formation. Le dernier exemple en date significatif à cet égard est encore observable dans les dispositifs d'insertion et de qualification mis en place pour les jeunes avant leur remplacement par le Crédit formation. Dans ces dispositifs de la "deuxième chance", le problème de l'évaluation a été soit scotomisé: absence de pratique évalutive dans les stages d'insertion, soit traité à la légère et de manière dévalorisante pour les individus par la création de sous-diplômes dévalorisés à l'avance car non reconnus par les entreprises -par exemple le C.F.G. (Certificat de Formation Générale)-, soit atomisé sous la forme de validition partielle dans le cadre des C.A.P.U.C. (Certificats d'aptitudes professionnelles par unités capitalisables) dans les rares cas où ils existent. Les jeunes ne s'y sont pas trompés et s'ils n'attendent plus rien sinon pas grand chose de ce dispositif au mieux ils l'utilisent comme solution d'attente de meilleures occasions. A l'oeuvre dès l'école maternelle -cf. les travaux de LURCAT- le processus de déscolarisation opère tout du long du cursus scolaire et de formation professionnelle.

Le second plan sur lequel on peut situer la crise de l'évaluation est le plan des techniques. En France, PIERON crée la docimologie en 1922. L'étude systématique des examens entreprise à cette époque a montré que les résultats produits par les pratiques d'évaluation présentaient de graves défauts de fidèlité et débouchaient sur des injustices flagrantes dans la mesure où le résultat d'un examen fait appel au hasard pour une large part. Malgré leur intérêt évident, les résultats de ces découvertes anciennes ne sont toujours pas prise en considération d'une manière consistante par les institutions éducatives. Il en résulte un retard technique important qui participe au manque de professionnalisation des enseignants sur le plan de la psychopédagogie. Acceptées, malgré leurs imperfections, comme moyens de sélection et de tri socio-professionnel les pratiques évaluatives restent négligées dans leur dimension pédagogique et formative. Certes des travaux sur l'évaluation formative sont menés, mais ils conservent largement un caractère expérimental et discret. Par ailleurs les essais limités de leur diffusion se sont soldés par des échecs tant sont preignants les habitudes et les réflexes qui accompagnent les formes classiques de l'évaluation.

Au changement radical ne doit-on pas préférer la politique des petits pas? Dans ce cas, une piste possible serait l'introduction et l'élargissement de l'utilisation des tests pédagométriques. Cette piste a été choisie par le Ministère de l'éducation à partir de la rentrée 1989/90. Les tests pédagométriques ne doivent pas être confondus avec les tests psychologiques qui visent une exploration de la personnalité considérée en elle-même. Les épreuves pédagométriques s'intéressent au domaine des acquisitions de connaissances, de savoirs et de savoir-faire scolaires, techniques et professionnels.

Dans ces conditions, la rédaction d'un ouvrage qui puisse faire le point sur la question de l'évaluation et présenter dans le détail ce qu'est un test pédagométrique nous a semblé utile comme participation positive à une évolution nécessaire des pratiques évaluatives.

L'ouvrage se divise en douze chapitres. Les trois premiers développent une présentation générale de ce qu'est l'évaluation, de ses fonctions et de son fonctionnement. Les quatre chapitres suivants proposent une analyse critique des pratiques évaluatives classiques en se plaçant successivement du point de la mesure, du point de vue des effets psychologiques et enfin du point de vue des effets sociaux de l'évaluation. Le huitième chapitre expose l'alternative d'une évaluation formative avant de passer à la question de l'utilisation des tests en pédagogie. Cette question est abordée en quatre volets. Nous expliquons d'abord ce que sont les tests dont le modèle nous vient de la psychologie et nous invitons à participer à la construction d'un test. Ensuite, les principes de l'implantation d'une épreuve seront abordés ainsi que les principes qui peuvent conduire le choix d'un test en réponse à des besoins particuliers.

Notre intention en écrivant ce livre se projette donc sur trois catégories d'objectifs. D'abord, faire le point sur la problématique de l'évaluation en pédagogie en reprenant l'ensemble des travaux les plus intéressants sur cette question; ensuite expliquer aux enseignants et aux formateurs ce qu'est un test pédagométrique; et pour finir donner les renseignements pratiques et les conseils utiles pour démarrer dans l'emploi des tests. Bien évidemment, nous ne ferons croire à personne que l'utilisation d'une épreuve d'évaluation standardisée va à elle seule permettre de surmonter la crise de l'évaluation évoquée à grands traits ci-dessus. Plus modestement nous pensons qu'en plus de la garantie d'objectivité scientifique qu'il apporte l'usage des tests oblige à un premier décentrage par rapport aux pratiques usuelles. Un premier décentrage en facilite un second et nous croyons que l'usage résolu des tests qui permettent analyse des erreurs et diagnostic pédagogique conduira progressivement les pratiques vers plus d'évaluation formative, c'est-à-dire vers de moins en moins de jugements de valeur formels et vers de plus en plus d'évaluation au service de la qualité des apprentissages individuels des élèves et des stagiaires. Aider chaque apprenant à situer ses connaissances d'une manière objective, mais aussi et surtout le stimuler dans la prise de conscience de ses démarches de résolution, de ses techniques de mémorisation, de ses processus de traitement de l'information, du rationnel de ses choix stratégiques... c'est au fond à la fois participer à un mouvement d'instrumentation des pratiques et à un mouvement de recherche active de la compréhension du fonctionnement individuel. Objectiver un niveau et comprendre le mode de fonctionnement qui "explique" ce niveau équivaut à se donner deux conditions hautement favorables à l'efficacité pédagogique.

CHAPITRE 1

QU'EST-CE QU'EVALUER EN PEDAGOGIE ?

L'évaluation en pédagogie porte sur les effets, sur les résultats concrets et observables d'une pratique d'enseignement théorique ou technologique, d'une action de formation générale ou technique.

Aucun enseignant, aucun formateur, aucun pédagogue, aucun psychologue n'est en mesure, en l'état actuel des connaissances, d'expliquer clairement et de manière indiscutable ce qui se passe réellement au cours du processus d'enseignement et de formation.

Nul ne saurait encore rendre compte avec précision de la nature exacte des effets de l'acte pédagogique. L'acte pédagogique est censé provoquer une transformation de l'élève ou du stagiaire. Cette transformation relève-t-elle d'un ordre physiologique, ou d'un ordre psycho-moteur, ou d'un ordre psychologique, ou d'un ordre cognitif, ou d'un ordre affectif, ou d'un ordre social, ou d'un ordre culturel... ou de tous ces ordres à la fois? A quelles conditions cette transformation devient-elle effective? Quels processus mobilise-t-elle?... De nombreuses années de recherche s'avèrent nécessaires pour apporter des réponses définitives si tant est que cette ambition puisse se réaliser.

Le degré d'incertitude relatif à la nature des effets de la formation des enfants, des adolescents et des adultes amène à penser que l'évaluation de ces effets est chose fort délicate.

Toutefois, face à notre propre ignorance, il convient d'avancer dans la compréhension de l'ensemble des facteurs et du jeu de leurs interactions qui interviennent dans les situations évaluatives. A cette condition, tout en gardant à l'esprit que nous restons toujours à un niveau très empirique, à la périphérie des

processus fondamentaux, nous pourrons accroître certainement d'un degré nos connaissances professionnelles et améliorer des pratiques d'évaluation qui s'appliqueront avec un plus haut niveau de conscience et par conséquence avec un plus haut niveau de maîtrise.

En première approche, on peut simplement envisager que l'évaluation en pédagogie n'est qu'une application parmi d'autres du sens général du terme. Le dictionnaire LAROUSSE, nous apprend qu'évaluer veut dire: "déterminer, apprécier la valeur, le prix de quelque chose, d'un bien, etc".

Le terme de valeur renvoie, toujours dans un sens commun, à de nombreuses notions. Ces notions prennent toutes une connotation à caractère économique, commercial, qui admet l'idée d'une négociation sociale.

La valeur d'une chose lorsque cette chose se définit comme un capital peut s'exprimer par son prix. En termes d'évaluation scolaire on peut se remémorer le prix d'excellence, la distribution des prix qui scandait la fin des années scolaires. Puisqu'il y a prix, il y a valeur d'échange symbolique, possibilité d'équivalence. Le titre scolaire peut s'échanger contre un emploi; l'équivalence admise de diplômes permet la reprise d'un cursus de formation dans un nouvel établissement... Définir la valeur d'une chose nécessite la mise en oeuvre d'une opération d'estimation telle que cotation, expertise, appel d'offres... Cette opération met souvent en jeu une mesure conventionnelle qui s'appuie sur des systèmes de signes ou de symboles. L'examen, le concours, les diverses formes d'interrogations écrites ou orales... constituent autant d'exemples d'opérations d'estimation qui utilisent elles aussi des systèmes de notes littérales ou chiffrées, de barêmes, de mentions...

La valeur peut exprimer le caractère de ce qui produit l'effet voulu. Dans ce sens on évoquera la valeur d'une méthode mais aussi la valeur d'un élève selon qu'il restitue les effets attendus ou non. La valeur exprime ce par quoi la personne est digne d'être estimée moralement, intellectuellement, professionnellement...

La valeur prend une tonalité particulièrement subjective lorsqu'elle exprime un attachement à des souvenirs personnels. En matière d'évaluation pédagogique la dimension subjective du jugement ou de l'interprétation de la mesure ne fera de doute pour personne. Elle s'alimentera au système des valeurs constitué de l'ensemble des unités de croyance que chacun pose comme vérités premières, système signifiant de l'idéal à atteindre.

Puisque le mot évaluation suggère autant de choses, on peut raisonnablement supposer que cela accompagne le fait que pratiquement toute activité humaine est régulée par le jeu continu d'un processus psychologique d'évaluation. L'activité de l'homme vise d'une manière générale l'atteinte d'un but qu'il s'est fixé, la réalisation d'un objectif qu'il a défini, la production d'un résultat dont il s'est construit une représentation... L'évaluation, qu'elle soit consciente ou inconsciente, qu'elle soit intuitive ou qu'elle use d'un dispositif méthodologique ou technique particulier, va consister à comparer les résultats effectifs de l'activité considérée aux résultats attendus. Cette comparaison évaluative opère tout du long du déroulement de l'activité afin d'en infléchir éventuellement le cours dans la

direction choisie: elle permet le "pilotage" de l'action ou du travail. Une activité de conduite d'un véhicule illustre convenablement l'intervention de ce type d'évaluation à finalité régulatrice. La comparaison opère également au terme de l'activité pour juger de l'écart entre le résultat attendu et le résultat obtenu, pour décider si l'objectif a été atteint, pour apprécier dans quelle mesure on s'est approché du but. La finalité de l'évaluation se déplace alors de la régulation vers le bilan de l'action ou du travail. Ce bilan prend la forme d'acquis expérentiels, d'acquis techniques, ou d'acquis scolaires selon les milieux où s'exerce le type d'activité. Ces acquis vont permettre, en conséquence de l'enrichissement qu'ils induisent, de mieux affiner le choix des buts futurs, de mieux différencier les moyens de leur atteinte et de mieux moduler la mise en oeuvre de ces moyens.

Bien que l'évaluation soit un processus d'accompagnement constant des conduites, qui participe à leur régulation et au bilan de leurs effets, il faut souligner que ce processus intervient à des niveaux de conscience et de technicité différents qu'il convient de distinguer.

L'évaluation non-dite, implicite ne fait l'objet ni d'un travail social ni d'une déclaration spécifiques. Processus psychologique inconscient, elle est présente à côté de toute activité. Il est possible d'inférer une part de sa logique par l'observation de ses effets sur les postures, sur les paroles et sur les conduites effectives. Par exemple les manifestations des attitudes liées à la réprobation commencent par le simple détour du regard. Le rejet s'exprime d'un revers de main. L'enseignant ne réagit pas explicitement à l'énoncé d'une réponse fausse mais interroge un autre élève ou repose sa question. Le public applaudit l'exploit du sportif. L'amateur de bons crus fait claquer sa langue ou tend une nouvelle fois son verre pour être de nouveau servi...

Difficile à aborder directement, l'évaluation implicite est d'autant moins à négliger qu'elle constitue la structure et qu'elle donne la direction des autres niveaux d'évaluation.

L'évaluation verbale courante consiste à émettre un avis, à formuler une opinion, à exprimer un jugement, à porter une appréciation, sur un objet, sur une personne, sur un événement... à l'occasion de libres conversations ou en réponse à une enquête d'opinion. Ces énoncés formulés de manière subjective sont monnaie courante dans de nombreux types d'échanges sociaux et ce dans toutes les strates de populations. Ils apparaissent tout aussi bien dans les dynamiques de recherche d'un partage, d'un concensus, que dans les dynamiques de recherche des oppositions, de provocation des conflits. Ils interviennent principalement dans l'instauration du climat émotionnel des groupes et dans la constitution de l'état affectif des personnes comme éléments d'existence, d'identité et de cohérence. Existence, identité et cohérnce de la personnalité seront remises en question, soit pour être renforcées, soit pour être ébranlées, à chaque opportunité de la production et à chaque occasion de la perception des énoncés évaluateurs.

Le jugement peut intervenir dans un cadre institutionnel. Que ce soit le cadre judiciaire, le cadre scolaire, le cadre économique, le cadre médical, le cadre scientifique... l'évaluation donne lieu à la mise en oeuvre d'un acte technique

inscrit dans une procédure repérable et répétable, procédure conduite par une personne socialement habilitée et reconnue.

Le recours à des outils, des instruments, des méthodes spécialement élaborées fait de l'évaluation institutionnelle l'affaire d'un professionnel qui soit capable d'expliciter selon les règles et les codes symboliques en usage dans le cadre considéré les résultats de l'acte d'évaluation.

En pédagogie, tout comme dans les autres domaines, l'évaluation est une notion très ample qui recouvre un ensemble complexe de techniques avec lesquelles on entend mesurer et/ou juger l'importance et la qualité des apprentissages.

Une conception étroite de l'évaluation est celle qui ne considère que les aspects instrumentaux, que les techniques de mesure et de notation, que les problèmes de mise en forme des résultats. A cette conception technicienne s'oppose une approche globalisante qui se nourrit des analyses du statut de l'évaluation dans le champ socio-politique. Réduire un phénomène à l'une ou à l'autre de ses extrémités induit la propre réduction de sa signification et en conséquence de sa compréhension. Une conception différenciée de l'évaluation prendra en compte l'ensemble de tous les facteurs qui interviennent dans le déroulement concret de sa mise en oeuvre: procédures, outils, méthodes...; processus de production du jugement évaluatif; caractéristiques des évaluateurs et des évalués, définition de leurs rapports spécifiques; articulations du contexte local aux contextes sociaux plus englobants...

Dans un cadre élargi, on remarque l'usage de nombreux concepts relatifs à l'évaluation. Deux termes sont quasi synonymes: estimation et appréciation. Que l'on parle d'évaluation, d'estimation, ou d'appréciation, il y a toujours élaboration d'un jugement de la valeur de l'objet considéré à partir de certains critères et dans le but de prendre une décision. Cette décision peut être un classement qui consiste à attribuer un rang à l'objet évalué, ou qui consiste à le rattacher à une catégorie. La décision peut être également de l'ordre de la sanction, de la validation des acquis, c'est-à-dire de l'ordre de la reconnaissance sociale d'une compétence, d'une capacité... Cette reconnaissance peut donner lieu à une certification, c'est-à-dire à la production d'un document officiel, titre ou diplôme.

L'élaboration du jugement évaluatif peut s'appuyer sur le résultat d'une notation qui utilise un système de lettres -A, B, C, D, E- ou de chiffres -échelles de 0 à 10; de 0 à 20- selon un barème pré-établi. Cette élaboration peut également s'étayer sur le produit d'une mesure effectuée avec un instrument spécifique -test pédagométrique- qui permet la comparaison à une unité ou à un système de grandeurs pris comme référence.

Le jugement de valeur participe d'autres fois à l'établissement d'un pronostic ou d'un diagnostic. Le pronostic est la conjecture faite d'après certaines informations de ce qui doit arriver d'une manière probable. Le pronostic intervient dans les actes de sélection, d'orientation, de passages de classes, d'admission dans un stage ou une école... La démarche de diagnostic est utilisée pour identifier la nature d'une situation, ou pour déterminer les caractéristiques d'un système à des fins de

maintenance ou d'amélioration. Le diagnostic alimente les rétro-actions qui prennent en compte les effets de l'action pour en réguler le cours.

Notons que le jugement de valeur peut être relatif ou absolu. Dans le premier cas, le jugement de l'élève ou de sa performance se fait par rapport à ses pairs, élèves d'un même âge, fréquentant un même cours...; la mesure se réfère à une norme, norme souvent statistique. Dans le second cas, l'évaluation est centrée sur l'apparition du répertoire comportemental attendu eu égard aux objectifs poursuivis; la mesure se réfère à un critère de maîtrise.

Reste à établir la différence entre évaluation et contrôle. Le contrôle se situe plutôt comme moyen d'obtenir des informations, des renseignements sur un fonctionnement institutionnel ou individuel, afin d'en rendre compte à une autorité supérieure et/ou éventuellement d'établir une comparaison en terme de conformité à un modèle de référence. Le contrôle, lorsqu'il s'opère en toute rigueur et dans la pureté de sa finalité, vise par essence à l'établissement de faits objectifs. Il ne permet pas les "états d'âme", et exclut toute interprétation. Le contrôle-type est présent dans les modèles de comportements théoriques proposés par les administrations civiles et militaires. L'établissement de la procédure et le respect de l'application de cette procédure l'emportent sur les autres aspects. Un contrôle des connaissances consiste uniquement à identifier de manière objective et indiscutable ce qui est connu par rapport à ce qui devrait l'être. L'évaluation dégage une perspective plus large puisqu'elle intègre un jugement de valeur qui peut se construire à partir d'une procédure de contrôle mais pas nécessairement, et puisqu'elle implique un questionnement sur le sens des faits objectivés et/ou des jugements formulés, en bref puisqu'elle propose une interprétation qui conditionne une compréhension suivie éventuellement d'une prise de décision.

Concept englobant, l'évaluation s'inclut elle-même comme moment de la démarche plus générale et plus complète qui conduit l'établissement d'un bilan.

Ample domaine de part le nombre de ces notions, l'évaluation se nuance et se diversifie d'une autre manière, par ses objets et par ses formes. L'évaluation, en pédagogie, peut porter sur le système de formation dans son ensemble ou sur une action de formation particulière; sur le formateur et sur les méthodes qu'il utilise ainsi que les objectifs qu'il poursuit; sur le groupe-classe et les résultats de l'apprentissage; sur l'individu, élève, étudiant ou stagiaire et de façon plus fine soit sur ses performances, soit sur ses compétences, soit sur ses maîtrises, soit sur ses capacités, soit sur ses connaissances, soit sur ses comportements, soit sur ses procédures de résolution, ses modes opératoires, ses méthodes d'exécution observables qu'il utilise, soit sur les processus et les représentations mobilisés pour la résolution d'un problème, l'exécution d'une tâche, la réalisation d'un travail inférables à partir de l'abstraction de la logique des comportements observables y compris la production des erreurs, soit, d'une manière plus générale sur un état de sa personnalité.

L'évaluation se tourne vers le passé au moment du bilan. Elle s'inscrit dans le présent au moment du contrôle continu des connaissances ou au moment du diagnostic de

la manière d'agir de l'apprenant; elle projette son produit dans l'avenir au moment du pronostic ou de la décision d'orientation...

Elle peut être concentrée sur l'une ou l'autre des phases du processus pédagogique ou sur tout ou partie de l'ensemble des facteurs qui interviennent dans ce processus. Elle se tourne enfin parfois sur elle-même et se pose comme son propre objet quand il s'agit d'évaluer ses résultats en regard de ses procédures.

Les formes de l'évaluation en pédagogie apparaissent multiples. La liste proposée ne prétend nullement à l'exhaustivité: interrogations orales et écrites dans toutes la diversité de leurs formes -exercices, "colles", problèmes, compositions, dissertations, devoirs, récitations, études, devoirs communs...-; examens académiques; tests pédagométriques; contrôle continu; examen partiel; concours; mais aussi conseils de classe et d'établissements, enquêtes et rapports administratifs ou de recherches appliquées.

La diversité des objets de l'évaluation et son polymorphisme expliquent sans doute en partie sa double valence affective. L'évaluation est perçue comme positive lorsqu'elle s'inscrit dans une dynamique de changement, de progrès, d'adaptation, de modernisation, de rationalisation des pratiques pédagogiques mises au service des élèves ou des stagiaires. A contrario, l'évaluation est perçue comme négative quand elle est utilisée à des fins de répression, de sélection aveugle, d'orientation imposée, de contrôle autoritaire...

Le dépassement de ce clivage devrait faciliter la remise en cause raisonnée de certains aspects de la formation, au niveau des formes d'organisation pédagogique, au niveau de la didactique des matières et au niveau de la relation éducative, vecteur de la transmission des apprentissages et des acquis et valeurs culturels.

Assurer ce dépassement nécessite de mieux comprendre à quoi sert l'évaluation en pédagogie.

CHAPITRE 2

A QUOI SERT L'EVALUATION EN PEDAGOGIE ?

Dans ce second chapitre, nous allons essayer de discerner et de mieux définir quelles sont les grandes fonctions servies par les actes, les procédures et les méthodes d'évaluation en usage dans les milieux éducatifs et scolaires. Des vues critiques sur l'exercice de ces grandes fonctions seront développées dans les chapitres quatre à huit. Dans le présent texte, l'objectif reste simplement un objectif de description.

Nous aborderons successivement les fonctions pédagogiques, les fonctions psychopédagogiques, les fonctions sociales, les fonctions psychosociales, les fonctions de recherche, les fonctions politiques de l'évaluation en pédagogie.

Ces différentes fonctions, nous les considèrerons toujours présentes, mêlées ensemble, dans tout acte d'évaluation pédagogique, qu'il appartienne à n'importe lequel des cinq grands moments où intervient l'évaluation.

Le premier de ces moments est celui de la passation des examens qui sanctionnent la fin d'un cursus scolaire par l'attribution ou non d'un diplôme d'enseignement général, le baccalauréat par exemple, ou d'un diplôme professionnel, le brevet d'enseignement professionnel - le B.E.P. - par exemple.

Le second moment est celui des épreuves écrites et/ou orales qui se déroulent durant la passation des concours d'entrée dans telle ou telle école normale, dans telle ou telle école d'ingénieurs..., ou dans telle ou telle administration. On appelle ces deux moments des évaluations externes dans la mesure où les évaluateurs ne sont pas les enseignants ou les formateurs habituels, et où les contenus des sujets proposés à la perspicacité des élèves sont définis par une autorité institutionnelle hiérarchique, c'est le cas des inspecteurs de l'éducation par

exemple, ou par une autorité de compétence comme une jury composé de chefs d'entreprises et de salariés d'un secteur professionnel. C'est également une autorité externe qui prend les décisions d'attribution ou de non-attribution du diplôme, d'admission ou de non-admission dans l'établissement scolaire ou dans l'administration...

Le troisième moment de l'évaluation pédagogique est celui où les élèves, les étudiants ou les stagiaires "planchent" sur des devoirs, des compositions dont les résultats vont alimenter les bulletins mensuels ou trimestriels. Les notes y figurent généralement accompagnées d'un commentaire écrit par les enseignants. L'ensemble de ces données sert en général de base de décision pour accèder à l'étape suivante du cursus suivi dans un même établissement. Elles s'intègrent dans le dossier scolaire utilisé parfois dans le cadre de l'attribution d'un diplôme, le brevet des collèges par exemple, ou dans le cadre d'un changement d'établissement, à l'occasion de l'entrée au lycée par exemple... De plus en plus fréquemment, il s'agit d'une évaluation "semi-externe" qui prend la forme de devoirs communs à l'échelon d'un niveau de cours d'un même établissement (toutes les classes de cinquième du même collège subissent la même épreuve d'anglais...), ou à l'échelon d'une circonscription (toutes les classes du cours moyen deuxième année de toutes les écoles primaires dépendantes d'une même inspection subissent des épreuves de français et de calcul identiques...). Evaluation semi-externe donc dans la mesure où le plus souvent l'enseignant va participer avec ses collègues à la défiition des contenus d'interrogation, à l'établissement d'un barème de cotation, à la correction des devoirs, à l'élaboration du commentaire évaluatif...

Le quatrième moment où s'exerce un acte d'évaluation pédagogique est celui de l'interrogation ou celui de l'exercice de contrôle des acquisitions scolaires, des apprentissages techniques, des maîtrises ponctuelles... La fréquence de "l'interro" ou de "l'exo" dans une même matière est à peu près hebdomadaire. Les questions portent en règle générale sur le contenu de la dernière leçon, sur le thème du dernier cours, sur l'objet de la dernière démonstration, sur le sujet du dernier exposé... C'est une évaluation interne dont le maître, le professeur, ou le formateur assurent, au moins dans le principe, l'entière responsabilité. Les résultats obtenus, quantitatifs et qualitatifs, peuvent alimenter le bulletin de notes de l'apprenant dans le cadre du contrôle continu des connaissances et donner lieu au calcul de moyennes.

Le cinquième moment ne se projette pas dans une phase ou une procédure d'évaluation formelle. Il s'inscrit en accompagnement de l'exercice courant de la conduite de la classe et s'intègre dans le fonctionnement habituel de la relation éducative. Nous retrouvons là le niveau de l'évaluation verbale évoqué au chapitre premier dans sa variante pédagogique. Cette évaluation verbale prend pour objet la performance ou le comportement du groupe classe considéré dans son ensemble ou la performance ou le comportement d'un élève en particulier. Elle va prendre forme et consistance dans un discours magistral composé de commentaires adressés soit à tout ou partie du groupe-classe soit à un élève précis. Ces commentaires transmettent soit des jugements évaluatifs positifs tels que des compliments, des félicitations, des encouragements..., soit des jugements négatifs tels que des remarques, des répri-mandes, des dépréciations... Plus discret encore, ce type d'évaluation permanente

rejoint l'implicite dans l'exercice de l'interaction sociale sous la forme des attitudes, des mimiques, et en général sous toutes les formes de "non-dit" expressif et signifiant.

Le détail du fonctionnement de l'évaluation à ces différents moments sera développé au chapitre suivant. Toutefois, dans la lecture de ce qui va suivre, c'est-à-dire dans la lecture de la description l'une après l'autre des différentes fonctions de l'évaluation, il convient d'essayer de garder à l'esprit, comme pour déployer une toile de fond, que toutes les fonctions décrites sont présentes et intimement liées à chacun de ces moments de l'évaluation. L'apparente simplicité d'une liste traitée terme après terme ne doit pas faire oublier que dans le domaine de l'évaluation nous sommes en présence de phénomènes psycho-sociaux complexes. Même si la signification de ces phénomènes ne peut se construire qu'étape par étape, il faut essayer de ne pas oublier pour autant sa globalité liée à la nature psycho-sociale de ces phénomènes.

En premier lieu, les pratiques spécifiques à l'évaluation remplissent deux types de fonctions pédagogiques. Le clivage entre ces deux types de fonctions est le statut du temps dans la procédure mise en oeuvre. Soit l'évaluation sert à faire le point à un moment donné: on prend la "photographie" des acquisitions à l'instant t; soit l'évaluation cherche à concrètiser ou à anticiper une évolution: on réalise le "film des acquisitions" effectives ou à venir entre deux instants t1 et t2. Dit autrement, une manière de classement des fonctions pédagogiques de l'évaluation procède par distinction des fonctions qui participent au bilan des acquisitions pédagogiques de celles qui sont au principe de la régulation de l'action pédagogique. Dans le premier cas, pour reprendre la terminologie proposée par SCRIVEN, on parlera plutôt d'évaluation sommative, dans le second cas, on évoquera plutôt une évaluation formative. Cette dictinction sera reprise et développée au chapitre huitième. Si l'on en reste à l'énoncé des fonctions assurées par l'évaluation sommative, on peut noter que celle-ci sert à appréhender la somme, le total des connaissances, des savoirs, des savoir-faire... acquis et maîtrisés par un élève et/ou un groupe d'élèves. Ce bilan constitue l'essentiel de la matière des examens et des concours, et aussi des devoirs mensuels ou trimestriels. Il peut servir aux administrations, selon les cas locale, départementale, régionale, ou nationale pour contrôler le rendement global du système éducatif et de formation, et pour vérifier que les grandes directives contenues dans les instructions et programmes officiels sont suivies dans un mouvement d'ensemble. Il sert également aux enseignants pour contrôler les acquisitions ponctuelles des élèves, pour savoir ce qu'ils ont appris et pour mesurer a posteriori l'écart entre les résultats atteints concrètement et les résultats attendus. Le bilan pédagogique formel doit permettre à l'enseignant, au formateur ou à l'examinateur de recueillir une information objective et fiable qui lui assure une base indiscutable pour poser son jugement de valeur autrement que sur des impressions vagues ou entièrement personnelles et subjectives. Dans un souci de plus grande objectivité, la performance de l'élève peut être comparée à des normes établies sur des échantillons représentatifs: c'est la méthode des tests pédagométriques qui sera expliquée au chapitre neuvième.

Quant à l'évaluation formative, elle n'intervient pas en fin de cycle de formation, mais de manière continue, durant tout le cours de l'année scolaire ou

pendant toute la durée de l'action de formation. Elle permet d'estimer la progression de chacun des élèves vers les objectifs visés par l'enseignant. Distinguant explicitement ce qui est acquis de ce qui ne l'est pas, l'évaluation formative apporte des éléments de modulation de l'action pédagogique par la modification éventuelle des contenus enseignés, par le changement des méthodes de transmission des savoirs, ou même par la restructuration de l'organisation pédagogique en constituant des groupes de niveaux par exemple. L'évaluation formative, parce qu'elle est une démarche continue d'observation et d'interprétation des effets induits par les pratiques d'enseignement, participe à la régulation et à la guidance générales de l'apprentissage de chacun des élèves. Elle alimente de faits significatifs les prises de décisions. Elle oriente l'intervention magistrale dans le sens du renforcement et de la cristallisation des acquis lorsqu'ils se font justement et dans le sens de la médiation corrective face au risque de perpétuation des erreurs.

Par rapport à la pédagogie vue comme l'art et la manière d'enseigner et de former, la psychopédagogie introduit les données relatives à la psychologie de l'enfant, de l'adolescent ou de l'adulte, et d'une manière plus générale les données relatives à la psychologie des apprentissages.

De ce point de vue, l'évaluation doit permettre à l'enseignant et au formateur de pouvoir dégager les facteurs favorables et les facteurs défavorables aux progrès de chacun et en conséquence de pouvoir différencier et individualiser leurs interventions en fonction des besoins perçus et identifiés.

Une aide précieuse pour l'ajustement des pratiques et des organisations consiste à connaître, grâce aux résultats fournis par l'évaluation, l'écart entre la nature des besoins manifestés par les élèves et celle que l'on peut abstraire des objectifs institutionnels de formation.

D'une manière plus fine, les démarches qui relèvent de ce qui est parfois dénommé évaluation diagnostique se centrent sur l'observation des procédures de travail de chaque élève afin de pouvoir expliquer le degré de recevabilité du résultat par la manière de procéder mise à l'oeuvre. Constater que le résultat d'une multiplication est faux est une chose utile qui permet de conclure à la non-maîtrise de ce type d'opération. Mettre en évidence pour l'analyser la démarche qui conduit l'élève à produire ce résultat erroné, connaître la source de son erreur: oubli des retenues, oubli du décalage au changement de ligne, méconnaissance d'une table, méconnaissance de la manipulation des virgules, cumul de difficultés traduisant l'incompréhension de la nature de la tâche, désintérêt total pour l'exercice, perturbation affective passagère,... est une autre chose nécessaire à la définition du contenu et de la modalité de l'intervention corrective. Cette intervention peut se référer à deux logiques principales. La première consiste à aménager, en fonction des données fournies par l'observation évaluatrice, des itinéraires singuliers les mieux à même d'amener chacun des élèves ou chacun des stagiaires à l'atteinte d'objectifs posés a priori et identiques pour tous, par application du principe que "tous les chemins mènent à Rome", ou encore du principe d'équifinalité si l'on préfère utiliser un langage savant. La seconde logique d'utilisation des données de l'évaluation procède par définition d'objectifs individuels, établis en fonction des points forts et des points faibles de chacun et se concrètisera dans des plans de travail personnels,

dans l'utilisation de l'enseignement assisté par ordinateur... Quelque soit le choix, que l'on préfère ajuster les objectifs ou que l'on préfère différencier les méthodes, il reste nécessaire de prendre en compte les données affectives à la source de bien des blocages. Dans le cadre de ce dernier aspect, notant que l'évaluation est d'un intérêt évident pour servir la régulation du dynamisme propre aux activités de l'apprentissage. La prise en compte des indicateurs comportementaux adéquats permet d'aider l'élève à mieux canaliser son investissement dans le travail, par exemple en l'incitant à prendre le temps de lire l'énoncé avant de commencer la résolution du problème s'il veut se donner une chance de le comprendre. De la même manière, l'évaluation sert pour le maintien des motivations à la condition de stimuler les progrès et d'éviter d'amplifier les régressions. A ce sujet, des expériences précises menées par LEPLAT et PETIT dans une formation à la mécanique générale ont montré que la connaissance par les élèves des résultats qu'ils obtiennent et des critères utilisés pour évaluer leur travail a exercé une influence significative sur la qualité des apprentissages.

On peut aller encore plus loin dans le sens du support que l'évaluation peut apporter à l'élève pour l'aider dans sa réflexion personnelle sur sa formation et pour l'aider à réguler ses apprentissages par lui-même, au minimum à participer à cette régulation. On aborde là les fonctions des pratiques de l'auto-évaluation ou de la co-évaluation. Ce type d'évaluation réflexive, dans la mesure où elle fait appel à la capacité de distanciation par rapport à sa production -qu'est-ce que je viens de produire? quelle est sa valeur?-, à la capacité de décentration par rapport à son propre fonctionnement - comment suis-je parvenu à ce résultat?-, à la capacité de projection dans l'espace social -quelle est la signification de ma production? de son degré de réussite?-, à la capacité d'une projection transcendantale dans le temps -ce travail satisfait quels besoins immédiats? quels besoins futurs?-... va induire chez le sujet apprenant des modifications psychologiques certaines dans la direction de son développement intellectuel, social et affectif, à travers l'adhésion à ce qui peut devenir alors son projet personnel de formation et donc de transformation admise et souhaitée de sa propre personne.

De la simple connaissance des acquis à la mise en place de la capacité d'auto-régulation des apprentissages dans la dynamique du développement des phases d'un projet individuel, on voit une grande diversité des fonctions psychopédagogiques de l'évaluation qui peuvent toutes se diriger dans le sens de l'aide aux élèves et dans le sens de l'amélioration générale du système éducatif et de formation.

Quelles sont les fonctions sociales de l'évaluation? Nous en distinguerons deux principales: la fonction de sélection et la fonction de certification. La sélection consiste à décider du choix parmi un certain nombre de candidats de ceux qui apparaissent comme les plus compétents pour tenir avec succès un emploi dans le monde du travail, ou pour aller au terme d'une formation générale ou d'une formation professionnelle. La sélection s'opère sur dossiers, sur concours, sur tests d'aptitudes et/ou de connaissances, sur entretiens, sur C.V., sur épreuves graphologiques (bien que ce type d'épreuves ne présente aucune valeur pronostique!)... La certification consiste à délivrer un document officiel reconnaissant un certain nombre de compétences générales, techniques, professionnelles...

Ces deux fonctions complémentaires trouvent leur raison dans les systèmes de définition des positions sociales et participent activitement à la distribution des individus sur ces positions.

La société est divisée en différentes strates socio-professionnelles. L'I.N.S.E.E. en propose une typologie en huit catégories principales :

0 – agriculteurs exploitants (8,5%) de la population);
1 – salariés agricoles (2,4%);
2 – patrons de l'industrie et du commerce (9,3%);
3 – professions libérales et cadres supérieurs (8,2%);
4 – cadres moyens (11,6%);
5 – employés (9,5%);
6 – ouvriers (45,9%);
7 – personnel de service (2%);
8 – autres catégories (2,6%).

L'accès aux multiples métiers que recouvrent ces différentes catégories se réalise selon des procédures plus ou moins complexes, dans des systèmes administratifs ou corporatistes ou familiaux plus ou moins ouverts. Mais dans tous les cas, il y est question à un moment ou à un autre de certification et/ou de sélection et donc d'évaluation. Ces catégories sont hiérarchisées ne serait-ce que sur le plan économique des salaires ou des revenus. Les systèmes éducatif et de formation participent, au moins d'une manière indirecte par la délivrance des diplômes qui sont pris en considération lors de l'attribution du statut social à l'entrée dans la vie active, au maintien de ces inégalités. Cela ouvre un débat technico-idéologique qui sera abordé au chapitre septième.

Il est vrai également que toutes les procédures, méthodes et outils d'évaluation formels et informels qui entrent en vigueur avant les examens ou les concours qui conditionnent le niveau d'entrée dans la vie active vont, dès l'émission verbale ou non-verbale du premier jugement de valeur en petite section d'école maternelle, préparer le résultat de l'épreuve de certification ou la décision de sélection finale, ainsi que les résultats des épreuves et des décisions intermédiaires.

Certification et sélection participent au contrôle social, législatif et règlementaire de l'accès aux emplois: elles apportent une garantie de la compétence de l'individu à assurer les responsabilités inhérentes au métier ou à la profession par la maîtrise prouvée et reconnue de ses opérations et de ses fonctions caractéristiques.

Egalement en jeu comme fonction sociale de l'évaluation qui opère en termes de certification et de sélection, la garantie que les candidats à un cursus de formation seront bien en mesure de le suivre à son terme et d'en tirer un profit optimal, le bénéfice individuel rejoignant alors la rentabilité globale du système éducatif et formatif.

Soulignons pour clore ce paragraphe sur les fonctions sociales de l'évaluation qu'évaluation, certification et sélection restent des processus autonomes, l'évaluation étant première, ses résultats alimentant les prises de décisions.

Les fonctions psycho-sociales de l'évaluation touchent plutôt aux procédures d'information et d'orientation et se réfèrent au problème général de la connaissance et de la reconnaissance sociale des individus et de leurs répercussions d'ordre psychologique.

Si les fonctions sociales précédentes de sélection et de certification ont un fonctionnement binaire, de type tout ou rien, au caractère décisionnel qui s'exprime par l'admission ou par le refus, les fonctions d'information et d'orientation apparaissent dans un système très nuancé qui par touches successives, va gérer l'élaboration du dessin des parcours de formation individuels et la définition des objectifs à atteindre. Elles procèderont par le conseil, la suggestion, l'entretien, l'apport et le commentaire de documents informatifs spécifiques tels que ceux distribués par l'O.N.I.S.E.P....

L'information et l'orientation sont assurées par les enseignants, les conseils et directions des établissements ou par un personnel spécialisé de psychologues scolaires et de conseillers.

Elles commencent dès l'école maternelle lorsque l'école demande soit un maintien en grande section, soit un passage anticipé au cours préparatoire, soit une admission dans l'enseignement spécial. Elle se continuent avec beaucoup plus d'ampleur au collège où une bonne moitié des élèves sera orientée vers la vie active sans qualification ou vers l'apprentissage précoce du métier au L.E.P. ou au C.F.A..

Ces fonctions combinent le jeu des libres entretiens "à la demande" entre les différents partenaires -enfants et élèves; parents; enseignants; directeurs; médecins scolaires; assistantes sociales; psychologues; conseillers d'information et d'orientation...-, avec les séances de devoirs de contrôle pédagogique ou de tests d'aptitudes ou de connaissances plus ou moins ritualisées, et la formalisation institutionalisée des jugements de valeur sur le bulletin scolaire, sur le dossier psychopédagogique, sur le dossier administratif...

Il peut en résulter une certaine ambiguïté, d'autant plus sensible que les décisions de certification et de sélection vont s'appuyer tout ou partie sur ces formalisations de jugements évaluatifs au caractère pronostic et sur les avis et conseils en matières d'itinéraires de formation.

Là encore il faut garder à l'esprit que les choses ne sont pas simples. Si l'on se doit d'informer les parents et les élèves sur le degré d'adéquation entre aspirations personnelles, résultats scolaires effectifs et contraintes de niveau de poursuite d'un cursus ou d'entrée dans un nouvel établissement, il faut bien considérer que ce type d'information n'est pas neutre et peut provoquer des effets psychologiques tout autant positifs, dans le cas de l'entretien d'un dynamisme orienté dans un projet reconnu, que négatifs lorsque l'information ou le conseil

brise toute illusion d'accès à une filière de formation, détruit tout espoir de réussite sans ouvrir d'alternative acceptable.

Les fonctions de recherche pédagogiques utilisent également des dispositifs d'évaluation divers –observation des activités spontanées ou des activités provoquées, observation libre ou armée; utilisation de moyens d'enregistrement de séquences didactiques; usage de tests, de questionnaires; analyse de statistiques... –.

Ces recherches peuvent essayer par exemple les moyens et les méthodes susceptibles d'harmoniser les besoins individuels et les objectifs institutionnels.

L'évaluation y sera utile pour apprécier et pour contrôler les effets d'une nouvelle méthode pédagogique, les avantages d'une modification de l'organisation des structures de l'établissement, les bénéfices tirés du remplacement de certains objectifs désuets par d'autres plus en rapport avec la modernité, les acquis liés à l'utilisation d'un nouveau logiciel...

Ce type de recherche peut être le fait d'un personnel spécialisé –chercheurs de l'I.N.R.P. par exemple–, de groupes associatifs –tel le G.F.E.N.... –, d'enseignants, de psychologues, de rééducateurs... ayant suivi des cursus de formation à la recherche dans le cadre de leur perfectionnement professionnel...

Enfin, les données de l'évaluation pédagogique peuvent servir de référence dans le cadre politique général du contrôle des dépenses que l'Etat consacre à l'éducation et à la formation.

Suite à ce cheminement fonction par fonction, rappelons notre idée force que pour pouvoir se dessiner une vue globale de l'utilité des pratiques d'évaluation, il convient de considérer que l'ensemble de ces fonctions interagissent à chaque circonstance de l'évaluation, de manière visible, manifeste, ou bien de manière plus diffuse, latente.

Si le jeu des fonctions évaluatives est relativement facile à discerner au moment d'un examen ou d'un concours officiels, il est tout aussi présent dans le simple cadre du commentaire verbal occasionnel énoncé par l'enseignant ou par le formateur. Dire à un élève ou à un stagiaire que sa réponse est correcte, c'est à n'en pas douter lui renvoyer le reflet de la valeur réelle de sa performance. C'est aussi lui faire prendre conscience, d'une manière induite, de sa propre valeur et si le commentaire positif se répète, c'est participer à la construction du sentiment de sa compétence. A partir de là, cet élève ou ce stagiaire risquent fort d'amplifier leurs investissements dans l'apprentissage et le cycle réponse correcte– gratification et reconnaissance sociales– augmentation du sentiment de compétence ira s'alimentant à la hausse de ses propres effets. Sauf incident de parcours imprévisible, les chances probables de réussite aux examens vont croître d'une manière directement proportionnelle à cette majoration dialectique. Les étapes de l'orientation et de la sélection seront facilement franchies. Le succès attendu va alimenter la reconduction des méthodes, des programmes, des organisations... qui l'ont permis. On pourrait,

bien sûr, développer la même description dans un sens négatif, à partir de l'accumulation des commentaires dévalorisants.

Comme le souligne si bien J.M. BARBIER, "... la fonction d'une pratique ne se réduit pas à sa finalité; si l'on veut approfondir la signification d'une pratique, il convient de faire l'inventaire de ses effets conscients ou inconscients, contrôlés et non contrôlés".

Afin de pouvoir réaliser cet inventaire, nous allons, dans le chapitre suivant, essayer de comprendre le fonctionnement des pratiques d'évaluation dans leurs différentes dimensions.

CHAPITRE 3

COMMENT FONCTIONNE L'EVALUATION EN PEDAGOGIE ?

La compréhension du fonctionnement interne d'un objet technique n'est pas une condition nécessaire à son utilisation. On peut conduire une voiture sans pour autant connaître le cycle du moteur à quatre temps ni le rôle de la bobine dans le circuit d'allumage... On peut se servir de l'ordinateur sans rien savoir du fonctionnement des éléments de l'unité centrale... D'ailleurs, la plupart du temps, pendant que nous faisons usage d'objets ou que nous nous livrons à différentes activités nous n'avons pas à l'esprit les processus fonctionnels à l'oeuvre et bien souvent nous ne connaissons qu'une partie de ces processus. Il est préférable qu'il en soit ainsi pour tous nos comportements automatisés. Ne dit-on pas qu'il suffise de réfléchir à comment on fait pour marcher pour commencer à trébucher? Combien pourrions-nous lire de mots à la minute si dans le temps de la lecture nous nous interrogions sur les processus de décodage en jeu?... De la sorte, on peut participer au déroulement de différentes procédures d'évaluation sans avoir une claire conscience de leurs mécanismes internes. La réflexion sur l'activité ne peut opérer pleinement et efficacement durant l'exercice de cette activité. Elle a meilleure place avant ou après, dans un espace-temps qui lui est consacré. Ce troisième chapitre propose donc cette réflexion sur le comment des pratiques d'évaluation.

Pour approcher cet objectif de description du fonctionnement de l'évaluation dans les milieux scolaires, éducatifs et formatifs, nous allons essayer de considérer l'ensemble des étapes d'une procédure d'évaluation sommative. Nous avons retenu ce type d'évaluation car il est de beaucoup le plus répandu et donc le plus familier. D'autre part l'évaluation formative retiendra longuement notre attention au chapitre huitième et nous reviendrons au sujet de l'évaluation comme processus intégré au fonctionnement de la relation éducative au chapitre sixième.

Dans ce chapitre encore nous nous tiendrons sagement à l'écart des discussions critiques, voire polémiques, qui régulièrement ressurgissent autour des problèmes que pose l'évaluation. Nous tenterons plus tranquillement de rechercher un ensemble de faits qui puissent rendre compte des différentes étapes d'une opération d'évaluation.

Onze étapes seront présentées dans l'ordre suivant: la définition des objectifs de l'évaluation; le choix des indicateurs; la construction de l'épreuve; l'élaboration du barème; la définition des conditions de la passation de l'épreuve; la correction; la notation; le classement; l'appréciation verbale; la synthèse des résultats et leur interprétation; la prise de décision.

Bien entendu, il s'agit là d'un découpage théorique qui n'évoquera pas les mêmes segments de réalité pour chacun des lecteurs. Plusieurs arguments justifient ce propos. Dans les faits, il est extrêmement rare qu'une seule personne ait à imaginer, concevoir, mettre au point, formaliser et appliquer l'ensemble des phases de la procédure complète. Par exemple, en règle générale, les personnes qui conçoivent les épreuves d'un examen ne sont pas celles qui les font passer ou qui les corrigent.

Dans les faits toujours, il arrive assez fréqemment que toutes les étapes ne soient pas explicitées ou même directement visibles. S'il est de tradition, dans le cadre d'une épreuve de dictée orthographique, de retirer deux points par erreur, l'élaboration de ce type de barème conventionnel ne sera pas explicitée à chaque fois.

L'ordre des étapes citées n'est pas immuable. Selon les cas, une étape peut se situer à un rang différent. Il arrive, par exemple, que le barème soit élaboré avant la passation de l'épreuve lorsque le critère retenu ne se réfère qu'à la connaissance de la matière évaluée, alors que d'autres fois il est élaboré après la correction afin de le construire de telle sorte que l'on puisse classer les candidats autour de la moyenne dans des proportions décidées à l'avance.

Il existe des procédures qui ne comportent pas l'ensemble des étapes proposées. A titre d'exemple, les professeurs et les formateurs n'écrivent pas un commentaire verbal explicatif de la note ou complémentaire de celle-ci par l'évocation de traits comportementaux à chacune de leurs interrogations.

La découpe du procès évaluatif dans les étapes proposées ne présente pas un caractère absolu. Il est certain que d'autres auteurs ont proposé, proposent et proposeront d'autres modes de partage. La justification de notre découpe est essentiellement une justification d'ordre didactique. Il fallait bien utiliser une liste d'étapes afin de faciliter la présentation organisée de la matière de l'exposé. Didactique, cette division n'en est pas pour autant arbitraire. Elle s'appuie sur l'observation des pratiques. Sa logique se justifie par la nécessité de saisir des unités fonctionnelles cohérentes lorsqu'on les considère prises isolèment, ayant donc un certain degré d'autonomie concrètisé par un fonctionnement propre. Ces unités, en sus, doivent pouvoir s'articuler entre elles-mêmes avec suffisamment de cohérence pour pouvoir décrire l'ensemble du processus évaluatif.

Soulignons enfin que les exemples présentés pour illustrer les propos proviennent de différentes procédures d'évaluation, dans différents contextes, dans différents moments... Nous espérons de la sorte élargir au maximum le champ des représentations relatives à l'évaluation. Cet élargissement devrait permettre en complément de sa diversité de dégager au travers de celle-ci une communauté conceptuelle constitutive de la pédagogie en général et de la pédagogie de l'évaluation en particulier.

Pour donner des résultats recevables, valables et compréhensibles, une procédure d'évaluation sommative doit offrir un certain nombre de garanties. Tout d'abord, l'évaluation doit répondre aux exigences communes à toutes les démarches qui englobent un processus de mesure et dont le prototype a été défini principalement dans le domaine des sciences physiques. Ces exigences seront reprises et développés dans le chapitre consacré aux qualités des tests. Ici, contentons-nous de les citer. La première exigence est celle de la validité, c'est-à-dire que le résultat produit doit être hautement représentatif de la chose évaluée. Selon la formule consacrée, la procédure d'évaluation doit évaluer ce qu'elle est censée évaluer. En second lieu, le résultat de l'évaluation doit être fidèle. Idéalement indépendant du moment et des circonstances de sa production, reproductible à volonté, il sera constant pour un même élève ou un même stagiaire. La troisième exigence est celle qui correspond au pouvoir de dicrimination de la procédure: les résultats obtenus doivent permettre de différencier les élèves selon leur niveau. A cela, ajoutons que pour être acceptable et acceptée par l'ensemble des partenaires intéressés, la procédure a un caractère public, c'est-à-dire un degré élévé de transparence. Dans le même but, ses résultats sont élaborés, consignés et transmis dans un langage compréhensible par tout un chacun.

1 - LA DEFINITION DES OBJECTIFS DE L'EVALUATION

En règle générale, le cadre de la définition des objectifs de l'évaluation relève de l'autorité scolaire. Les objectifs appartiennent aux textes et aux instructions officielles émanant du Ministère. Chacun des étages administratifs et hiérarchiques intermédiaires entre le Ministère et les enseignants peut, à l'intérieur du cadre et par le biais de lettres-circulaires, préciser des contenus particuliers et les codifier de manière plus ou moins rigide et plus ou moins détaillée. Notons aussi le rôle non-négligeable de la tradition orale qui se transmet entre les générations d'enseignants. Concrètement, cela se traduit par le fait que le degré de liberté de l'enseignant est tout relatif dans le domaine de l'évaluation. Ce degré sera quasi nul dans les épreuves de certification gérées sur un plan national comme le baccalauréat par exemple et au cours desquelles l'enseignant est invité à appliquer le plus rigoureusement possible les directives qui lui sont communiquées. Il sera plus important à l'occasion des interrogations de contrôle qu'il organise dans sa propre classe, au moins en ce qui concerne le choix des sujets parmi l'ensemble du programme imposé. Les élèves et les stagiaires ne sont pas partie prenante dans la définition des objectifs de l'évaluation.

A l'occasion de la définition des objectifs, il est utile de distinguer entre l'objet et la finalité de l'évaluation.

Les objets possibles apparaissent nombreux et diversifiés. Contentons-nous de n'en citer que quelques-uns: mettre en évidence les progrès de chaque élève ou les progrès de l'ensemble d'un groupe-classe par rapport à un corpus de questions de référence de manière à pouvoir objectiver les étapes et l'amplitude des acquisitions réalisées; dans le cadre d'une recherche pédagogique ou dans le cadre de l'application des directives accompagnant la mise en place d'une réforme, montrer les effets de tel cours ou de telle méthode; au terme d'un cursus de formation, évaluer un savoir ou un savoir-faire final; chercher à mettre en évidence un trait psychologique, une capacité, une disposition comportementale qui ait une valeur prédictive particulière...

Le travail d'évaluation à partir de ces objectifs peut s'orienter vers plusieurs finalités différentes, qui elles aussi ont intérêt à être clairement définies, comme l'ajustement et la régulation des pratiques d'enseignement ou de formation; le contrôle ponctuel ou continu des connaissances; la certification des acquis de l'enseignement général, de l'enseignement technique ou de la formation professionnelle; le conseil ou la décision d'orientation scolaire ou professionnelle; la sélection de candidats à un cursus ou un stage de formation... Nous ne développerons point ces finalités largement évoquées dans le chapitre deuxième relatif aux fonctions de l'évaluation en pédagogie.

Ce moment de définition préalable est décisif quant à la fixation du degré de différenciation des objets et des finalités de l'évaluation. Dans l'idéal, il ne devrait rester quasiment aucune confusion comme, par exemple, entre l'évaluation des caractéristiques d'une réalité et cette réalité elle-même. Il est préférable, au titre de la transparence des intentions, de savoir et de dire si l'évaluation va porter sur les contenus académiques restitués, ou sur la manière dont ils sont formulés et présentés, ou sur les procédures de résolution et les comportements développés à l'occasion de la réalisation des tâches, ou sur des qualités psychologiques qui seront inférées de ces comportements et procédures de résolution... juge-t-on la copie ou l'élève, la production, la manière de produire ou le producteur?

2 – LE CHOIX DES INDICATEURS

Imaginons que l'objectif soit l'évaluation des acquis en langue anglaise des élèves d'une classe de quatrième, dans le cadre du bilan de fin d'année.

La première chose à considérer est la définition de l'univers de contenu à évaluer. Ici, le contenu global de l'évaluation est représenté par l'ensemble des connaissances en langue anglaise tel qu'il est précisé dans le programme officiel des classes de quatrième des collèges. Cet ensemble constitue lui-même un sous-ensemble de la langue anglaise en général.

Dans les faits, il peut arriver que le professeur ne puisse "boucler" son programme. Dans ce cas, le contenu prend en considération les éléments de connaissances effectivement enseignées, les éléments d'expérience réellement vécues par les élèves. L'univers du contenu de l'évaluation présente donc un caractère

public et il se situe d'abord comme extérieur à l'élève. La procédure d'évaluation doit donc mettre en évidence dans quelle mesure les contenus enseignés ont été intériorisés par l'élève, dans leurs aspects "passifs" (compréhension de la langue) et dans leurs aspects "actifs" (production autonome de la langue). C'est le fait que l'intériorisation des contenus ne se concrètise pas de la même manière en quantité et en qualité chez les différents élèves qui justifie la nécessité d'une évaluation. Dès la troisième année d'apprentissage de la langue anglaise il est tout à fait impossible de réaliser l'inventaire exhaustif des connaissances d'un élève particulier et a fortiori d'une classe entière. Il s'avère vite nécessaire de procéder par sondage. Cette contrainte pose le problème du double choix et de la double délimitation des domaines partiels du domaine général à évaluer et des contenus propres à chacun de ces domaines partiels. Que choisir? le vocabulaire ou la conjugaison? le thème ou la version? la vie familiale ou la vie sportive?... Et si on choisit la vie sportive anglaise, faut-il se centrer sur les sports traditionnels, sur les sports aristocratiques, sur les sports populaires, sur l'histoire des sports, sur la pratique du sport à l'école...

Les domaines partiels et les contenus spécifiques retenus vont prendre valeur d'indicateurs dans deux directions. Ils constituent en premier lieu un échantillon des contenus les plus significatifs d'un domaine de connaissances. En ce sens, ils doivent être représentatifs du domaine à évaluer du fait qu'ils en couvrent l'ensemble des capacités sans toutefois les manifester toutes. Lacunaires par rapport à la totalité du domaine évalué, il n'empêche qu'ils n'en portent que l'essentiel. Dans l'autre sens, les indicateurs sélectionnés sont pertinents quand réellement ils permettent une inférence sur le contenu effectivement assimilé par l'élève. Au fond, la fonction de l'indicateur consiste à rendre possible l'opération de comparaison entre la structure et le contenu du domaine académique et la structure et le contenu du domaine psychologique. La valeur virtuelle de l'indicateur se perçoit en conséquence comme directement proportionnelle à son degré d'efficacité d'opérateur de comparaison épistémologique. Il doit être représentatif du comment est construite la connaissance et du comment se construit la connaissance.

Cependant, ce n'est pas parce qu'il possède ce pouvoir qu'il faille inverser les niveaux: contrairement au propos que l'on prête à A. BINET, "l'intelligence est ce que mesure mon test", l'indicateur n'est pas ce qu'il indique, mais seulement un informateur. On n'évalue pas l'indicateur, mais ce sur quoi il renseigne, du moins dans le temps de l'évaluation. On peut, en d'autres circonstances, s'interroger sur la valeur propre de l'indicateur à partir de la quantité et de la qualité de l'information qu'il permet d'obtenir dans le souci du perfectionnement technique des outils de l'évaluation.

3 – LA CONSTRUCTION DE L'EPREUVE

Une fois les indicateurs sélectionnés et définis, il convient de passer à la phase d'élaboration de l'épreuve. Cette élaboration touche au contenu et à la forme. L'épreuve fonctionne comme un stimulateur de comportements, de conduites, de réponses qui puissent s'interpréter en termes de maîtrise d'un domaine de compétence. Au fond, l'épreuve s'interpose entre le savoir académique constitué et le savoir psychologique

en voie de constitution. Son mode de fonctionnement relève d'une double transitivité: si on admet, par construction de l'épreuve, qu'il y a correspondance entre contenu académique et réponse attendue par l'évaluateur; si on admet, par inférence du jeu intermédiaire d'une capacité de répondre, qu'il y a correspondance ente la réponse réelle et le contenu assimilé par l'élève ou le stagiaire; alors il y aura correspondance entre le contenu académique et le contenu assimilé quand la réponse réelle et la réponse attendue seront jugées identiques, interchangeables.

L'épreuve correctement élaborée sera vue, en dérivé de ce fonctionnement transitif, comme celle qui offrira le plus haut degré de probabilité de provoquer le comportement, la conduite, ou la réponse attendue. Plus ce degré de probabilité diminue, plus on se dirige vers le système des devinettes, et plus les réponses jugées conformes au standard attendu seront le fruit du hasard. L'accroissement du nombre de réponses possibles à une même question affecte l'ensemble des qualités de la procédure d'évaluation. L'augmentation du poids du hasard dans la stimulation de la production de la réponse discrédite la validité de l'épreuve qui ne mesure plus uniquement ce qu'elle doit mesurer. La perte de fidèlité est une perte obligée en ce sens qu'un même élève pourra fort bien produire des réponses différentes à une même question non de sa propre initiative mais du fait de l'imprécision, du flou, de l'ambiguïté... de l'épreuve. Si validité et fidèlité faiblissent alors la discrimination des élèves ou des candidats ne reflètera que la logique des rangs distribués au hasard et l'épreuve perdra en final toute signification.

Le contenu de l'épreuve est directement référé à l'indicateur. Si la définition de ce dernier est cohérente par rapport au domaine évalué, et suffisamment explicite et détaillée, alors le contenu en découle "automatiquement", au titre de l'évidence logique.

C'est de la qualité de la mise en forme de l'épreuve que va dépendre le degré de probabilité de la réponse attendue.

Dans un devoir d'histoire, si on demande "Que pensez-vous de Louis XIV?", il est clair que l'on obtiendra une plus grande diversité de réponses que si on demande "Quelles sont les dates du règne de Louis XIV?". L'objet qu'il sera possible d'évaluer dans le premier cas ne sera pas tant un contenu historique que la capacité de formuler et de justifier un jugement, une prise de position personnelle par rapport à un oeuvre politique. Ce type de question ouverte posera également un important problème pour décider de la recevabilité des réponses. En effet, il est certain que chaque professeur d'histoire ayant son propre jugement sur le règne de Louis XIV appréciera les réponses en fonction de ses propres opinions et le processus d'évaluation va se pervertir dans la recherche d'une certaine conformité subjective.

Le second type de question, ou question fermée, a pour objet une connaissance précise, admise d'une manière universelle. La restitution de cette connaissance met en oeuvre essentiellement une capacité mnémonique et relève de ce qu'il est d'usage d'appeler "l'encyclopédisme". Mais le but de l'éducation ou de la formation se limite-t-il à la simple restitution de connaissances universelles?

Un autre type de questionnement consiste à proposer plusieurs réponses possibles à l'élève ou au candidat qui doit choisir la réponse correcte: ce sont les Q.C.M., questionnaires à choix multiple. Par ce système, si la réponse produite par le candidat est judicieusement choisie, on est sûr à cent pour cent qu'elle correspond à la réponse attendue qui, par ailleurs est unique par construction. La simple mémoire, mais aussi toute autre capacité peuvent se trouver sollicitées par le Q.C.M.. Malgré tout, il reste toujours un doute sur la possibilité d'une réponse au hasard. Par exemple, sur un choix parmi quatre propositions il existe une chance sur quatre de réussir par hasard. On reproche au Q.C.M. de proposer des réponses fausses qui risquent d'être assimilées par association à la question dans le cas de leur choix par l'élève qui répond au hasard. A un niveau intermédiaire, le Q.C.M. facilite la production de réponses plutôt intuitives, c'est-à-dire des réponses dont le processus d'élaboration n'est que peu ou pas du tout développé. Il arrive également que l'on puisse répondre sur des indices partiels.

Soit la question: "Quel est le résultat de la multiplication suivante: 3 214 X 52
10 000 ? 42 ? 766 893 ? 167 128 ?"

Il suffit de considérer que le résultat se terminera par le chiffre 8 pour opérer le bon choix sans avoir de calcul à effectuer. De la sorte, on ne sait pas s'il faut référer la réponse correcte à un processus de tirage au sort, ou aux capacités relatives aux stratégies de choix, ou à la mise en oeuvre effective d'une capacité multiplicative, en principe seul objet de l'évaluation.

On voit, à travers l'analyse de ces trois formes de questionnement, toute la difficulté d'arriver à des formes irréprochables, c'est-à-dire à des formes susceptibles de provoquer une seule réponse attendue, réponse dont l'occurence puisse s'interpréter de manière univoque quant à la capacité psychologique mise en oeuvre par l'élève ou le candidat, dans son rapport au domaine de compétence évalué.

Il existe d'autres manières de poser les questions. Citons encore l'exemple de "l'exercice à trous", ou épreuve de closure qui fait appel à la capacité de complètement pour éprouver le domaine de compétence évalué . Ce type de forme est très adaptée quand on s'intéresse à l'utilisation des capacités psychologiques à l'ouvrage dans un domaine de contenu.

Soit l'exemple d'exercice de français suivant :
"Complète la phrase suivante: ... elle n'est ... trop violente la pluie arrose ... poireaux de manière ... et économique".

La résolution fait franchement appel aux processus d'inférence cognitive: il faut construire logiquement les propositions manquantes à partir de l'analyse des propositions données.

La lecture du "n'" donné amène à tester les deux mots possibles pour satisfaire la règle de la double négation: "ni" et "pas". L'absence d'alternative élimine la possibilité "ni". Il reste "pas" qui fonctionne correctement dans la phrase. Le "x" à la fin du mot "poireaux" laisse à penser que le déterminant ou l'adjectif numéral choisi pour être placé devant ce nom sera au pluriel. Le premier mot sera choisi pour

compléter l'expression de la condition amorcée par "trop". Que ce soit "Si" ou "Quand" ou "Comme" ce premier mot doit prendre une majuscule du fait de sa position en début de phrase. Le dernier "trou" appelle un adjectif susceptible de qualifier la manière d'arroser de la pluie. C'est la richesse du vocabulaire et la capacité de sa mobilisation qui intervient ici.

Cette forme de questionnement met donc en jeu un véritable complexe de connaissances de la langue française, de capacités à mettre en relation et à opérer sur ces connaissances, de capacités d'inférence logico-verbale. Si la réussite permet de conclure positivement sur l'ensemble de ces qualités, l'échec ne peut être d'emblée relié à l'une d'entre elles. La détermination de sa source, si on souhaite la connaître, passe par une investigation spécifique du fonctionnement individuel. De toutes les façons, on remarque une fois encore que la distinction des contenus académiques assimilés par l'élève des capacités intellectuelles sollicitées par la forme du questionnement choisi s'avère impossible.

Une fois la forme générale de questionnement déterminée, il reste à regarder les qualités, les caractéristiques rédactionnelles de chacune des questions posées, ou de la consigne relative aux tâches dont la réalisation est demandée. Si le concepteur de l'épreuve souhaite que la probabilité d'apparition de la réponse attendue soit la plus élévée possible, il doit poser les questions d'une manière explicite.

Soit la question de physique suivante : "Le principe d'Archimède".

Le candidat peut-il se contenter du simple énoncé du principe, y ajouter sa démonstration, développer quelques applications, raconter les circonstances de sa découverte et évoquer le célèbre "eurêka"...?

En résumé, la conception de l'épreuve dans son contenu, dans la forme de questionnement et dans la rédaction des questions ou des consignes doit surmonter de nombreux obstacles pour que les réponses produites soient directement interprétables par rapport aux objectifs posés.

4 – L'ELABORATION DU BAREME

Dans une évaluation sommative, le barème assure le passage entre le monde des contenus pédagogiques et le monde de la mesure. Il fait aller de l'aspect qualitatif des réponses de l'élève à l'aspect métrique d'une échelle de notes. Il fait passer du "juste ou faux" au "il a la moyenne, il n'a pas la moyenne, il a au-dessus de la moyenne". Il met en correspondance les réponses attendues et une valeur numérique. Il inscrit le produit pédagogique brut de l'épreuve sur une échelle de valeur qui en est tout à fait extérieure et indépendante, de la même manière que lorsqu'on construit le prix d'un article ou d'un service, en projetant sa qualité sur une échelle monétaire. Le barème fonctionne comme un taux de change bancaire ou comme un interprèteur en informatique: c'est un "logiciel" intermédiaire avec lequel on peut passer directement d'un langage à un autre, terme à terme.

Le barème agit à deux niveaux. En premier lieu, il distribue les points disponibles entre les grandes parties et leurs subdivisions. Ainsi, pour l'épreuve de chimie on notera les questions de cours sur 8 points et les exercices sur 12 points. S'il y a cinq exercices, les trois premiers peuvent être notés sur 2 points et les deux derniers sur 3 points. En second lieu, le barème va distribuer les points à l'intérieur de chacune des subdivisions du devoir en attribuant une certaine valeur à chaque réponse attendue. De la sorte, le barème va relativiser le poids des contenus sollicités et des capacités nécessaires à leur maniement. Cette répartition a priori doit refléter l'importance relative des parties de l'interrogation d'une manière qui soit cohérente avec les objectifs de l'évaluation. D'autres critères de pondération peuvent entrer en ligne de compte, comme le degré de complexité de chacune des questions. Plus un exercice nécessite de manipuler un nombre important d'informations, ou plus les conduites de détour nécessaires pour parvenir au résultat final y seront longues, ou plus le niveau d'abstraction y est élevé et mieux il sera "rétribué". De même, des notions-clés dans un programme pourront voir leur valeur majorée.

La fixation des seuils d'attribution des points en regard à des niveaux de performances, si elle ne présente pas un degré de réalisme suffisant basé sur la connaissance des populations examinées, risque de déboucher sur des résultats inacceptables qui nécessitent une révision du barème a posteriori.

5 - LA DEFINITION DES CONDITIONS DE LA PASSATION

Au cours de cette phase préparatoire, il s'agit de prendre les décisions relatives aux paramètres périphériques à l'évaluation proprement dite. Le terme périphérique n'est pas à prendre dans le sens de secondaire, mais dans le sens d'une caractérisation de ce qui entoure le centre. En effet, décider d'une durée d'une heure pour un devoir de composition française ou décider d'une durée de quatre heures pour le même devoir aura forcément des conséquences importantes sur le produit fourni par les élèves et sur les critères de correction et de notation.

En plus de la durée de l'épreuve, il faut décider de sa modalité: sera-ce une interrogation écrite, ou orale, ou un exercice à l'atelier...?

Est-ce une épreuve individuelle ou une épreuve de groupe? L'usage d'une calculatrice électronique de poche est-il autorisé? Ou celui du dictionnaire? Est-il permis d'apporter ses notes de cours ou d'utiliser une documentation personnelle? Quel matériel ou quel document seront fournis par l'administration et quels autres seront apportés par le candidat ou l'élève?

Les feuilles de brouillon devront-elles être de différentes couleurs selon les rangées d'examinés? Qui surveillera le déroulement de l'épreuve? Qu'est-il prévu de faire lorsqu'une fraude est constatée? Quels types de réponses pourront être donnés aux éventuelles questions des examinés? D'ailleurs, ont-ils la possibilité de poser des questions?

A quel endroit se déroulera l'épreuve? Dans la classe ou l'atelier habituels? Dans un autre établissement de la ville? Dans une autre ville?

Y a-t-il maintien du groupe d'élèves ou mixage des différentes classes, des différents établissements?

Qui va corriger l'épreuve? Le maître ou le professeur habituels ou d'autres enseignants? Y a-t-il constitution d'un jury et si oui qui le compose? Des enseignants, ou des formateurs, ou des représentaux syndicaux, ou des chefs d'entreprises...?

De quelle manière, sur quel support, sous quelle forme, dans quel délai, par qui, à qui,... les résultats seront-ils communiqués?

Bien entendu, toute ces prises de décision dépendront du cadre dans lequel se déroule la procédure et de la fonction de l'évaluation. Les conditions de la passation du concours de l'agrégation diffèreront des conditions de la passation de l'examen du permis de conduire!

6 - LA CORRECTION

Une fois l'épreuve passée, il sera temps de la corriger. Nous distinguons la correction de la notation: afin d'attribuer la note, il convient d'examiner les réponses fournies par les élèves ou les candidats et de distinguer celles qui sont recevables de celles qui ne le sont pas, de déterminer celles qui correspondent aux réponses attendues et celles qui n'y correspondent pas. La correction est donc le moment où l'on juge de la qualité de la réponse indépendemment de la notation. A l'exception de réponses très précises et très ponctuelles comme les dates historiques, les formules chimiques, des résultats numériques particuliers... en règle générale la réponse produite ne correspond exactement à la réponse attendue que rarement. Se pose alors la question du degré de tolérance, du degré d'amplitude de l'écart supportable pour que la réponse produite, par rapport à sa forme et/ou par rapport à son fond, soit admise comme assimilable, ou comme équivalente à la réponse attendue.

Le correcteur exerce là son jugement personnel, ce qui explique les différences éventuelles des appréciations portées par des correcteurs différents sur une même copie (cf. chapitre suivant), ou sur un exercice professionnel... Le système de la double correction, qui intervient lorsque l'enjeu social de l'évaluation est important, apporte une garantie plus forte d'objectivité sous la forme d'une synthèse de deux points de vue. Cette zone de libre interprétation de la réponse, si elle apparaît plus réduite, n'en existe pas moins même dans les cas de l'utilisation de corrigés-types. Par exemple, à un problème d'arithmétique, l'élève conduit correctement son raisonnement et ses calculs, mais dans l'expression de sa réponse il oublie d'indiquer l'unité. Est-ce admissible ou non admissible? Elle s'élargit quand on touche aux domaines littéraires et philosophiques où l'opinion et l'expérience personnelles, même étayées solidement par les connaissances universelles, vont commander le jugement.

7 - LA NOTATION

La note se construit en mettant en correspondance chacune des réponses correctes et un certain nombre de points. Cette mise en correspondance s'établit en appliquant le barème lorsqu'il existe. Le total des points attribués à chaque réponse est ensuite calculé dans les systèmes qui procèdent par sommation. Il existe d'autres systèmes qui procèdent par pénalisation et diminution du capital de points. C'est le cas de la notation des dictées où ce ne sont pas les réponses attendues qui sont comptabilisées, mais les erreurs. Un nombre de points est retiré pour chaque erreur, selon un code plus ou moins précis pouvant aller au quart de point. Dans le cas du "Cinq fautes, 0" ce système ne permet plus de différencier les élèves qui ont commis entre cinq et plusieurs dizaines de fautes.

L'une des difficultés de la notation réside dans le respect du centrage de l'évaluation, dans sa "pureté". Il est parfois délicat de ne pas mélanger les domaines. Par exemple, le notateur en mathématiques doit-il ou ne doit-il pas pénaliser les erreurs d'orthographe, ou les ratures?

La note rend compte d'un niveau de performance à travers un nombre ou une lettre: échelles de 0 à 10, de 0 à 20, de A à E... Plusieurs notes peuvent être remplacées par une note de synthèse, la moyenne. Ce niveau de performance est détaché de la compétence proprement dite. Avoir la moyenne en mathématiques ne renseigne pas sur les savoirs réellement connus et les savoirs-faire effectivement maîtrisés. C'est une indication de niveau scolaire, ce niveau en question étant une entité relativement abstraite représentative d'une sorte d'idéal à atteindre. Avoir 15/20 en anglais ne dit pas si l'individu peut écouter et comprendre la B.B.C., lire du Shakespeare en version originale, traduire un roman policier, demander son chemin à Londres, réciter toutes les formes des verbes irréguliers... Avoir 15/20, c'est connaître les trois-quarts du maximum que l'on peut connaître. La note, parce qu'elle appartient au domaine de la mesure, quantifie un volume de connaissances, mais dans le même temps elle éloigne de la connaissance de ces connaissances. De ce fait, elle peut prendre un aspect mythique, et ce d'autant plus qu'elle renvoie à la notion de norme. Par exemple, dans les cursus scolaires il est convenable d'avoir sa moyenne, c'est-à-dire d'avoir assimilé la moitié du programme. Ceci est tellement vrai qu'au niveau d'une classe des notes trop élevées, ou trop basses, ou trop dispersées vont déclencher des mouvements de protestation des différents acteurs. Un enseignant qui conduit bien sa classe est un enseignant qui sait tenir la moyenne. S'il attribue des notes trop élevées il sera suspecté de laxisme; s'il attribue des notes trop basses, il sera accusé d'extrême sévérité sinon d'incompétence.

L'aspect mythique de la note se trouve également renforcé du fait qu'elle devient rapidement jugement de valeur, non plus ou non seulement des acquisitions, mais du travail de l'élève, si ce n'est de l'élève lui-même: "l'élève Untel vaut 12 en sciences physiques!" Détachée de l'oeuvre et donc devenue indépendante du domaine des contenus de l'enseignement académique ou technique, la note se projette dans l'espace social et jauge d'une manière indifférenciée le producteur de cette oeuvre.

8 - LE CLASSEMENT

On passe de la note au classement pour situer la position d'une performance individuelle dans l'ensemble des performances d'une population de référence. Pratiquement on ordonne les notes puis on leur attribue un rang.

Cette population de référence peut être le groupe-classe ou la section de stagiaires, ou la promotion d'élèves d'une grande école... La population de référence peut être constituée d'une manière plus large, par exemple en regroupant tous les élèves d'un même cours dans un établissement, au niveau d'une ville, au niveau d'un canton ou d'un département... Dans le cadre des concours, c'est l'ensemble des postulants qui seront classés. Pour les tests pédagométriques, la population de référence, classe d'âge ou classe scolaire, est, par construction, représentée par un échantillon.

En dehors des concours de recrutement, le classement est de moins en moins formalisé au rang près. Malgré tout, on remarque assez souvent des indications concernant la note moyenne du groupe, la note la plus haute et la note la plus basse qui permettent de situer une note particulière par rapport à la note moyenne.

Entre classement et mérite il n'y a qu'un pas et le principe de base de la "méritocratie" postule la correspondance entre le classement scolaire et le niveau de l'intégration dans le corps social.

9 - L'APPRECIATION VERBALE

L'évaluation scolaire contient donc des éléments quantitatifs, exprimés à l'aide de lettres (notation littérale), à l'aide de nombres cardinaux (les notes) et à l'aide de nombres ordinaux (les classements). Ces éléments chiffrés de l'appréciation sont très souvent accompagnés d'un commentaire évaluatif verbal. Dans l'évaluation en pédagogie coexistent donc deux systèmes d'information.

Le commentaire verbal, lorsqu'il accompagne la note, remplit plusieurs fonctions. Tout d'abord, il va rapprocher la note de la production effective: en différenciant les acquis et les lacunes, il justifie la note mais en sus il peut proposer des objectifs de révision, de reprise de certaines notions,... Il peut proposer également une critique de la teneur des réponses et évoquer des alternatives d'idées, de développements, de formes de raisonnement, de modes de résolutions, de types d'organisation... Le commentaire verbal relativise la note attribuée en la situant par rapport aux performances précédentes obtenues par le même élève: une note de 12/20 n'aura pas une signification univoque pour un élève qui est passé de 18 à 12 ou pour un autre élève qui lui est passé de 8 à 12. On considère dans ce cas l'évaluation du point de vue de l'ipséité de l'élève.

Le commentaire relativise également la performance par rapport aux capacités perçues chez l'élève, par rapport à la somme de travail et aux efforts qu'il a fournis. Il peut rapprocher la qualité du devoir de qualités psychologiques ou de traits comportementaux de l'élève.

Il relativise enfin la note par rapport au degré de difficulté de l'exercice.

Le commentaire d'un ensemble d'épreuves qui évaluent le travail d'un trimestre par exemple en bas du bulletin scolaire attribue assez souvent un jugement de valeur relatif à la personnalité de l'élève: "élève sérieux, élève méritant,... , élève dissipé, élève paresseux..."; il peut également comparer le bilan d'ensemble au potentiel perçu intuitivement: "peut mieux faire, peut travailler plus, ne donne pas sa mesure... des progrès inattendus..."

L'appréciation verbale intervient indépendamment d'une notation à différentes occasions, mais rarement dans le cadre de l'évaluation sommative.

10 – LA SYNTHESE DES RESULTATS ET LEUR INTERPRETATION

Elaborer la synthèse des résultats, c'est ajuster, formaliser et expliciter le jugement de valeur porté sur la performance et/ou son auteur; c'est construire du sens, c'est attribuer une signification à cette performance, en la comparant aux éléments du contexte de référence, en la rapprochant des objectifs pédagogiques. La synthèse prend une tonalité positive ou une tonalité négative selon que l'évaluateur estime ce qui est acquis ou ce qui reste ignoré.

On peut considérer l'écart de la note à la norme chiffrée (moyenne, ou seuil statistique, ou note éliminatoire...), mais aussi l'écart en termes d'acquis minimum, ou d'acquis spécifiés par un programme, ou d'acquis majoritairement connus du groupe... On peut aussi analyser l'écart en termes de comportements, de valeurs...

Si la question est un indicateur du contenu académique, la réponse est elle-même un indicateur des connaissances de l'élève et de sa capacité à les utiliser, au moins dans un contexte évaluatif. L'interprétation portera donc sur le volume supposé des acquis scolaires effectifs et sur la capacité psychologique sous-jacente. Si l'on parvient à ce niveau d'interprétation s'impose la prise en compte des caractéristiques du sujet, des composantes de sa personnalité, des événements de son histoire... de la manière la plus consciente et la plus explicite possible. La compréhension globale de l'élève ou de l'apprenant va surgir de la confrontation des résultats objectivés par l'épreuve d'évaluation et des éléments de connaissance scolaires et extra-scolaires de cet élève. C'est bien cette compéhension-là qui marque le premier pas vers une pratique pédagogique plus nuancée en fonction de chacun.

L'interprétation des résultats, d'un autre point de vue, peut tenter de rapprocher le niveau de performance moyen atteint par le groupe et les conditions de l'apprentissage en général, la méthode pédagogique employée en particulier. On ne peut s'attendre à des résultats équivalents lorsqu'on a trente-huit élèves ou vingt-trois élèves dans sa classe...

L'interprétation peut enfin présenter un caractère projectif. Du résultat actuel on peut prévoir des comportements futurs, on peut anticiper de prochaines réussites, on peut pronostiquer l'adaptation à de nouveaux contextes scolaires ou

professionnels. Cet aspect projectif de l'interprétation entre en jeu dans la prise de décision.

11 – LA PRISE DE DECISION

La prise de décision peut intervenir dans tous les domaines fonctionnels de l'évaluation: informer l'élève, les parents, l'administration; certifier un niveau de compétences; orienter, sélectionner, recruter; modifier un programme ou une méthode, un matériel didactique...

La difficulté de cette prise de décision est liée à l'importance de ses conséquences sur la trajectoire sociale de l'élève ou du stagiaire, sur le devenir professionnel du candidat et à l'importance de ses conséquences sur l'équilibre psychologique des individus. La décision s'accompagne d'une responsabilité de nature déontologique.

Cette difficulté globale, majorée affectivement par l'ampleur des enjeux psychologiques, sociaux et professionnels, se rattache pour une part aux difficultés intrinsèques à l'évaluation. C'est d'abord le manque relatif de liberté de l'évaluateur qui à la limite risque de l'enfermer dans des procédures ritualisées qui s'auto-finalisent: on donne un devoir parce que l'administration demande n notes par trimestre et uniquement pour cette raison.

Les épreuves présentent-elles toujours des qualités métriques suffisantes qui garantissent leur fiabilité? Les indicateurs choisis sont-ils suffisamment pertinents pour pouvoir conclure sans risquer d'erreur d'appréciation grossière? Y a-t-il un nombre correct d'éléments qualitatifs qui puissent contrevalider d'une part et expliquer d'autre part la performance chiffrée et permettre de relativiser le jugement?

Sur le plan de l'interprétation, le choix du référent idéologique reste le point central: faut-il considérer que le niveau de performance objectivé se rattache à un don naturel, à une aptitude génétiquement déterminée par rapport auxquels toute volonté d'action relève de l'illusion, ou faut-il considérer que ce niveau est le produit des circonstances actuelles et qu'il peut changer en réaction à la modification de la conjoncture locale? Les décisions seront différentes si l'on croit que le niveau de performance est une donnée stable dans le temps ou si l'on croit qu'il peut s'améliorer.

C'est parce que l'évaluation est un domaine techniquement difficile et socialement sensible que ses pratiques sont la cible de nombreuses critiques.

CHAPITRE 4

QUE NOUS APPREND LA DOCIMOLOGIE ?

Les études docimologiques ont pour objet la validité des notes scolaires et des notes d'examen. Le terme docimologie combine deux mots grecs. Le premier, "dokimé" signifie épreuve et le second, "logos", signifie science. Ce terme fut proposé par Henri PIERON pour désigner l'étude systématique des examens.

On trouve dans le *Vocabulaire de la psychologie* rédigé sous la direction de ce même auteur la définition suivante :
"... étude systématique des examens (modes de notation, variabilité interindividuelle et intra – individuelle des examinateurs, facteurs subjectifs, etc.)".

On peut donc dire que la docimologie propose une critique scientifique des procédures de l'évaluation traditionnelle. Nous verrons au chapitre suivant que la docimologie s'occupe également de rechercher et d'essayer des solutions aux questions d'ordre technique que pose l'évaluation.

On a pu distinguer deux grands aspects dans l'étude des procédures d'évaluation. Tout d'abord l'aspect purement technique de l'examen, dans sa dimension métrique et dans ses qualités de validité et de fidèlité. Les méthodes d'étude employées dans ce domaine appartiennent essentiellement à l'ordre de la statistique. Cette approche centrée uniquement sur la technique porte le nom savant de docimastique.

La seconde approche, beaucoup plus large, va s'intéresser au rôle que joue l'évaluation dans la dynamique éducative, à ses effets psychologiques et à ses conséquences émotionnelles sur les élèves,... Elle emploie les méthodes de la psychologie, observations, tests, questionnaires, entretiens... On donne à cette approche le nom de doxologie.

La docimologie va donc s'occuper de recueillir tout un ensemble de données d'observation sur la manière dont les enseignants ou les examinateurs notent (données quantitatives) ou apprécient (données qualitatives) un devoir écrit ou oral, ou toute autre forme d'épreuve.

Dans un deuxième temps, la docimologie va se centrer sur les écarts éventuellement constatés. Elle va d'abord décrire ces écarts, tester s'ils sont significatifs sur le plan statistique, c'est-à-dire tester si on peut les imputer au hasard ou à un facteur qui jouerait systématiquement. Puis le chercheur en docimologie va tenter d'élucider quelles sont les causes des écarts de notation ou d'appréciation.

Les écarts en question sont observés dans deux types de schémas. Le premier schéma recherche les différences entre les notes attribuées à un même devoir par plusieurs correcteurs. Le second schéma prend en considération les notes mises par le même correcteur à une même copie mais à des moments différents.

Le troisième temps, une fois la probabilité causale établie, sera consacré à l'interprétation de l'ensemble des résultats de l'étude. La discussion des constats pourra même prendre un tour philosophique par le biais d'un questionnement sur la fiabilité, sur la justesse et sur la justice des examens.

Enfin, dans une dernière phase du déroulement de son cours, la démarche de l'étude docimologique peut éventuellement aller jusqu'à la proposition de méthodes d'évaluation qui soient plus rigoureuses car plus proches de l'objectivité.

Les travaux docimologiques et leurs conclusions ne sont guère connus du monde enseignant. Peut-être parce qu'ils se caractérisent par une tonalité critique et de prime abord peu constructive, tonalité susceptible de provoquer des réactions de rejet de la part des personnes qui mettent en oeuvre ces pratiques et ce d'autant plus que ces personnes seraient moins disposées à y changer quoique ce soit. Une autre hypothèse explicative de cette faible diffusion consiste à prendre en considération le fait que la docimologie fut élaborée dans la sphère des psychologues, sphère qui communique peu avec celle des enseignants du fait que les psychologues ne sont pas sollicités pour participer à la formation professionnelle des enseignants et réciproquement. Peut-être aussi parce que les psychologues qui opèrent sur le terrain de l'éducation et de la formation se cantonnent ou sont cantonnés hors du champ pédagogique en général et hors du champ de l'évaluation pédagogique en particulier.

Pour la petite histoire, notons que, comme bien d'autres découvertes, la docimologie doit sa naissance à un fait de hasard. En 1922 le couple PIERON assisté de LAUGIER mènent une recherche relative à la validité des tests utilisés à l'époque dans le cadre de l'orientation scolaire et professionnelle dont par ailleurs PIERON assure la responsabilité nationale. Ces chercheurs comparent les résultats obtenus par un même groupe d'élèves aux tests et les résultats obtenus à la session du C.E.P. du mois de juin 1922. Ils constatent que le degré de correspondance entre les deux types de résultats est très, très faible. Cette remarque amena les chercheurs à déplacer leur centre d'intérêt vers la valeur proprement dite des examens.

A partir de là toute une série d'études furent menées sur les examens d'abord, puis élargies à la doxologie. Lorsqu'on regarde l'ensemble de ces études, on conclut à une convergence quasi-absolue des résultats qui pointent tous un défaut d'objectivité des procédures traditionnelles de l'évaluation.

Toutes les conclusions des études docimologiques vont dans un sens identique: elles montrent des divergences sensibles et importantes entre les enseignants correcteurs.

A titre d'exemples, les pages suivantes exposent quelques-uns des résultats des études docimologiques.

En 1931, la Carnégie Corporation, fondation américaine, finance une vaste enquête internationale sur les examens et les concours. En France, ce fut sur le baccalauréat de l'année 1932 que porta l'enquête conduite par LAUGIER et WEINBERG.

PIERON décrit le dispositif adopté: "Il fut prélevé dans les archives de l'office du baccalauréat de l'Académie de Paris, par les soins de son directeur, M. Piobetta, 100 copies tirées au hasard d'une composition française, d'une version latine, d'une composition de langue anglaise et d'une composition de mathématiques, pour la 1re partie, d'une dissertation philosophique pour la 2e partie philosophie et d'une composition de physique pour la 2e partie mathématiques - les copies-recopiées - dont les notes avaient été relevées, furent confiées à six groupes de cinq examinateurs de baccalauréat, qui les notèrent et fournirent ensuite un rapport sur leur méthode propre de cotations, les qualités recherchées et leur hiérarchie, l'appel éventuel à des composantes distinctes, etc. Enfin, trois copies de composition française furent soumises dans les mêmes conditions à 76 correcteurs".

Deux tableaux résument les principaux résultats de l'enquête. Le premier tableau prend en compte les décisions d'admission à l'examen en fonction des notes attribuées par les six correcteurs.

Candidats / Disciplines	Admis par les six correcteurs	Refusés par les six correcteurs	Admis ou refusés
Mathématique	20 %	44 %	36 %
Anglais	16 %	37 %	47 %
Version latine	10 %	40 %	50 %
Physique	13 %	37 %	50 %
Compos. franç.	9 %	21 %	70 %
Philosophie	10 %	9 %	81 %

La colonne "admis ou refusés" est la plus éloquente: elle indique le pourcentage des désaccords au niveau de la prise de décision. On remarque, même pour les mathématiques réputées science exacte, un tiers de décisions discordantes. Le taux de désaccord prend une allure vertigineuse pour la philosophie: huit fois sur dix il y a désaccord entre les correcteurs !

De tels résultats montrent clairement la part de hasard qui entre en jeu dans le résultat à un examen. Et quand on pense au poids social que chacun accorde au baccalauréat, la question de la justice des décisions se pose d'elle-même.

Au niveau des notes attribuées (notes sur 20), il a été observé que l'écart le plus fréquent entre les correcteurs considérés par paire varie de quatre à sept points. Cet écart est le plus important pour la composition française et la philosophie et il est moindre pour l'anglais, les mathématiques et la physique.

A une même copie, la même note n'a jamais été attribuée et souvent une note est supérieure à la moyenne (10 points) et l'autre lui est inférieure.

Chacun des correcteurs utilise une échelle de notes qui lui est personnelle. Ces échelles varient d'une matière à l'autre de deux manières: les notes extrêmes ne sont pas les mêmes, et la moyenne des notes distribuées fluctue. Par exemple, selon les correcteurs les notes moyennes vont de 6,32 à 10 en composition française, de 7,11 à 9,48 en physique, de 7,65 à 11,23 en philosophie...

Une analyse de la variance (méthode d'analyse statistique qui permet d'étudier le "poids" respectif, l'importance relative des différentes sources de variation des notes) donne comme conclusion que les différences entre les candidats interviennent pour 40% et celles entre les examinateurs pour 60% dans la variation des notes. En clair, si l'on voulait pronostiquer la note, mieux vaut connaître les habitudes de l'examinateur que le niveau du candidat !

Pour les trois copies de composition française, les notes attribuées par les 76 correcteurs vont de 1 à 13 pour la première, de 3 à 16 pour la seconde, de 4 à 14 pour la troisième.

Le tableau ci-dessous donne la répartition des notes pour la première copie.

Notes	Nombre de correcteurs
0 - 1	1
2 - 3	6
4 - 5	20
6 - 7	34
8 - 9	10
10 - 11	3
12 - 13	2

Finalement, la conclusion de ces travaux est que la notation effectuée par un seul correcteur se rapproche du tirage d'une note au hasard.

Pour arriver à attribuer une "note vraie" aux épreuves du baccalauréat, les chercheurs ont calculé qu'il faudrait multiplier le nombre des correcteurs et réaliser la moyenne des notes que chacun d'entre eux attribueraient à la même copie.

Ainsi il faudrait treize correcteurs pour les mathématiques, seize pour la physique, dix-neuf pour la version latine, vingt-huit pour l'anglais, soixante dix-huit pour la composition française et cent vingt-sept pour la philosophie.

Citons plus rapidement quelques autres études. En 1934, trente-sept copies de certificat d'études supérieures de sciences sont notées une fois par trois professeurs différents et à deux reprises par deux autres professeurs. Cinq notes sont donc disponibles pour chaque copie. La note d'admission est 10/20.

Si 27% des candidats sont admis 5 fois et 19% ne le sont jamais, pour 54% d'entre eux la décision est fluctuante en fonction du correcteur. (LAUGIER, PIERON, PIERON, TOULOUSE, WEINBERG).

En 1955, une recherche prend comme objet la comparaison des moyennes des notes attribuées par dix-sept jurys de la série philosophie et de treize jurys de la série mathématiques du baccalauréat, jurys entre lesquels les candidats d'une même série étaient répartis au hasard, par ordre alphabétique. A l'écrit des mathématiques, les moyennes varient de 5,81 à 9,06 et pour la philosophie de 8,2 à 9,5. A l'oral de physique (série mathématiques) la variation est de 8,3 à 13 et à celui de sciences naturelles (série philosophie) de 9 à 14...

Les pourcentages d'admissibilité s'étalent, selon les jurys, de 48 à 61% en philosophie et de 31 à 53% en mathématiques. Avant même de passer les épreuves, leurs résultats sont donc en partie prédéterminés par les habitudes de notation des correcteurs et par la composition des jurys.

Il a été également observé de grandes différences de résultats dans l'évaluation d'un même élève, dans une même matière, entre l'interrogation écrite et l'interrogation orale (PIERON, REUCHLIN, BACHER, DEMANGEON).

On pouvait espérer que dans d'autres situations, sans enjeu social aussi fort que dans les circonstances de la passation d'un examen comme le baccalauréat, et dans un système d'évaluation plus qualitatif, c'est-à-dire d'une part moins rigide qu'une échelle de notes et d'autre part plus subtil, les fluctuations évaluatives se révèlassent moins impressionnantes.

Il n'en est rien. Par exemple, le C.R.D.P. de Lyon organise l'expérience suivante en 1969: on demande à 150 professeurs de mathématiques de classe de troisième de corriger et de noter sur vingt points la copie de trois élèves. Voici le tableau récapitulatif des notes attribuées.

	Note la plus basse	Note la plus haute	Note moyenne
Elève A	0,5	11,5	5,7
Elève B	11,5	20	16
Elève C	3,5	11,5	8

Résultats impressionnants qui n'ont guère besoin de commentaires !

DEMANGEON et LARCEBEAU en 1958 mènent une enquête auprès de cent trente quatre élèves de C.M.2 à partir d'une appréciation verbale sur une épreuve de calcul et une de français en quatre catégories: très bon, bon, moyen, faible. Chaque épreuve est évaluée par huit correcteurs. L'utilisation de la catégorie "très bon", pour le calcul varie de 15 à 23% et en français de 0 à 30%. Donc, dans le cas de l'utilisation d'une échelle qualitative, les références sur lesquelles s'étaye le jugement sont loin de présenter un caractère universel. Il en résulte un taux de discordance important dans la catégorisation des résultats. Relativement à une même copie, la catégorie choisie concorde dans 42% des cas en calcul et dans 16% en français; deux catégories adjacentes sont utilisées dans 48% des cas pour le calcul et dans 64% pour le français et trois catégories différentes dans 8 et 20% des cas repectivement dans chacune des deux matières.

La forme entièrement orale de l'interrogation amène des constats identiques. Par exemple, en 1962, PIERON, REUCHLIN et BACHER proposent à l'écoute de seize professeurs de lycée l'audition d'une vingtaine d'interrogations orales de physique. Les moyennes des notes fournies par chacun des correcteurs, moyennes qui se devraient d'être égales dans la mesure où chaque correcteur note les mêmes réponses et en supposant que la mesure du niveau soit fidèle, s'étalent de 8,03 à 13,40. Dans le contexte de l'évaluation pédagogique, les déterminants subjectifs de la note ou de l'appréciation ont donc la part belle !

Considérons maintenant le cas où l'enseignant est sollicité pour produire un jugement global, sans recours ni à la mise en place d'une épreuve spécifique, ni à une technique de notation précise. On trouve une étude de la S.R.E.S.A.S. qui porte sur le signalement d'enfants en difficulté parmi une cohorte de 1238 enfants fréquentant des grandes sections d'écoles maternelles et l'année suivante des cours préparatoires. Il fut demandé, successivement aux maîtresses de ces deux niveaux scolaires de "signaler les enfants posant des problèmes d'adaptation, ainsi que les raisons précises de ces signalements".

En grande section, la fréquence des signalements varie de 8,10% à 43,42% selon les écoles. La variation des fréquences relatives aux différents types de difficultés n'est pas négligeable: 20 à 60% pour les troubles du comportement, 2 à 36% pour les troubles du langage...

D'une année sur l'autre, il est établi que "... sur 189 enfants signalés en grande section, 73 seulement (soit 38,6%) sont signalés à nouveau au début du cours préparatoire..." et que "... 61 enfants qui n'avaient pas été signalés en grande section d'école maternelle sont signalés au cours préparatoire..."

Plus modestement, à l'échelon d'un groupe scolaire, une enquête sur les demandes de soutien en lecture du cours préparatoire au cours élémentaire met en relief un même phénomène de discordance des jugements magistraux: sur une cohorte de 135 élèves qui passent d'un cours à l'autre, un soutien est demandé pour 30 enfants parmi lesquels seulement 12 (40%) sont signalés à la fois en fois de C.P. et en début de C.E.1. (BALLION).

Qu'il soit exprimé d'une manière quantitative ou d'une manière qualitative, et qu'il soit élaboré par n'importe quelle procédure, le jugement magistral se caractérise par son aspect aléatoire et par sa faible valeur pronostique. Les procédures d'évaluation traditionnelles se présentent après analyse comme peu valides, peu fidèles, peu fiables. Pourtant elles sont à la base de nombreuses prises de décisions qui déterminent pour une part le devenir social et professionnel des élèves.

Tout cet ensemble d'études a pu montrer l'existence d'une triple variabilité des notes scolaires.

Si l'on confie un même lot de copies dans le même temps à un groupe de correcteurs qui travaillent sans se consulter, d'un correcteur à l'autre on obtiendra des différences significatives entre les moyennes des notes ainsi qu'entre leurs dispersions.

Si l'on donne un même lot de copies à un seul correcteur mais à des moments espacés dans le temps, on observe également des fluctations non négligeables.

Enfin, un élève n'aura pas des niveaux de performance identiques d'une situation d'examen à une autre.

Cette variabilité s'explique par une série de facteurs qui sont liés aux caractéristiques et aux habitudes du correcteur, aux contenus que propose l'épreuve et à la forme de leur présentation, aux caractéristiques de la population examinée.

Il est apparu que chaque évaluateur utilise sa propre échelle de notation: la notation est plus ou moins "sévère" en fonction de son niveau d'exigence qui dépend, en autres facteurs, du degré de maîtrise de la matière qu'il enseigne, des objectifs qu'il poursuit, de son expérience des diverses populations scolaires... L'étendue de l'échelle de notes dépend aussi des usages: mettre une note maximale à une copie de composition française (notée habituellement entre six et quatorze), cela ne se fait pas alors que c'est possible pour une copie de mathématiques (notée par usage de zéro à vingt). Il entre donc une très grande part de subjectivité dans la notation, et par delà la réponse effective, le correcteur réagit selon son propre goût et selon ses propres conceptions de l'enseignement, selon également sa compréhension personnelle de la question posée et de la réponse proposée. Il réagit selon des normes, des

habitudes, des stéréotypies plus ou moins personnelles, en référence à des modèles plus ou moins explicités. Les effets de halo, qui font que le jugement se généralise à partir de l'utilisation d'un seul critère, orientent et biaisent le résultat.

L'ordre des copies joue également par effet de lassitude lorsque le nombre de copies est important, par effet de contraste selon que la copie précédemment corrigée était bonne ou mauvaise.

Les connaissances ou les renseignements que le correcteur peut avoir de chacun des élèves, tels ses performances antérieures, ou son statut scolaire, ou ses origines socio-économique et ethnique, influencent son jugement. Par exemple, NOIZET et CAVERNI montrent expérimentalement que: "Les mêmes copies sont évaluées différemment lorsqu'on donne aux évaluateurs des informations différentes sur les élèves censés les avoir rédigées". D'après ces auteurs, les informations données a priori vont provoquer des attentes chez le correcteur par le détour de l'activation des représentations relatives à des types ou des catégories d'élèves. Ces attentes vont orienter la recherche et la perception des indices afin de ne pas placer le correcteur face à de trop fortes contradictions.

Le contenu de l'épreuve est forcément ponctuel et il ne peut avoir une représentativité que limitée de l'ensemble des connaissances à connaître ou des capacités à maîtriser. Sa mise en forme peut révéler des difficultés de communication entre l'auteur du sujet, l'examinateur et l'examiné tels que de possibles malentendus relatifs au vocabulaire utilisé, au choix des formulations...

L'élève ou le stagiaire peuvent avoir "la chance" de tomber sur un sujet qu'ils connaissent parfaitement ou être interrogés sur un point de programme ou de progression qu'ils n'ont jamais réussi à assimiler. La rencontre entre le candidat et le contenu de l'épreuve, rencontre tant redoutée, présente un caractère largement hasardeux. Selon que l'aléa est plus ou moins favorable, l'état de stress sera plus ou moins facilement surmonté. On sait aussi la force et la qualité de l'attention varie chez une même personne aux différents moments de la journée...

CHAPITRE 5

QUE NOUS PROPOSE LA DOCIMOLOGIE ?

A ses débuts, la docimologie a pris une tonalité plutôt négative du fait que les auteurs se préoccupaient essentiellement du constat objectif et scientifiquement établi du manque de fidèlité et de validité des procédures d'évaluation usuelles.

Ultérieurement, les chercheurs se soucièrent, dans un état d'esprit constructif, de formuler des propositions techniques susceptibles de diminuer l'ampleur des erreurs d'évaluation par l'amélioration des qualités métriques des dispositifs.

De la sorte on peut signaler que le courant de recherches docimologiques a inversé son cours en conséquence d'un déplacement de ses intentions. Après avoir concentré ses efforts sur le décryptage et l'objectivation de la réalité des pratiques et des résultats de l'évaluation scolaire, il quitte cette perspective essentiellement critique pour se poser en force de propositions, potentiellement capable de modifier cette réalité. Après avoir constaté le mal, des chercheurs universitaires conseillent, prescrivent un certain nombre de remèdes.

Par exemple, BONBOIR définit une docimologie positive, au service de la réussite scolaire, qui relie fortement évaluation et action pédagogique :
"... tout travail doit être apprécié... dans un souci réellement éducatif... le premier défaut, la première insuffisance, la première imperfection dans les acquisitions (doivent être repérés) chez les écoliers, afin que l'on puisse agir en connaissance de cause, avant qu'intervienne l'inévitable et l'irréparable échec... l'examen n'est pas une réalité parallèle à l'action pédagogique et accompagnant inévitablement cette dernière, mais il lui est intimement intégré...; ... évaluation "continue" et... "évaluation formative" mettent en cause le système d'enseignement tout entier".

Une première série de propositions est relative à l'amélioration des examens externes, c'est-à-dire aux examens qui dépassent le cadre de l'établissement et au cours desquels les enseignants ne rédigent ni ne corrigent les épreuves que passent leurs propres élèves. Le baccalauréat est un examen externe type.

On pourrait, pour réduire les divergences inter-correcteurs et inter-jurys dans l'attribution des notes, opérer quelques corrections statistiques simples consistant, par exemple, à égaliser les moyennes et les dispersions. De la sorte, moyenne et écarts à la moyenne seraient identiques pour tous les correcteurs.

Ce type de mesure soulève de fait deux objections. La première est d'ordre psychologique: on peut supposer qu'aucun correcteur accepterait l'idée que l'on puisse modifier les notes qu'il attribue "en son âme et conscience". La suivante est d'ordre technique: on ne peut, en toute rigueur, égaliser des moyennes qu'à la condition d'avoir constaté que les différents groupes d'élèves soient de niveaux sensiblement équivalents. Comment s'en assurer alors que les procédures d'évaluation habituelles ne sont ni fidèles, ni valides? Peut-on se permettre une pré-évaluation à l'examen? Cela aurait-il un sens social acceptable par tous ?

Un second train de propositions touche à l'organisation collective de l'examen. En premier lieu, l'administration de l'examen peut constituer les jurys de manière équilibrée, en prenant en compte le degré de sévérité des correcteurs. De la sorte, les écarts globaux inter-jurys, en termes de décisions d'admission, se retrouveront réduits, mais la question des écarts inter-correcteurs d'une même matière restera en suspens. Cette question est d'importance quand les résultats du candidat sont analysés dans une finalité d'orientation dans un type de cursus. Elle peut s'aborder à travers l'établissement d'un barême analytique de correction commun qui facilite l'adoptation d'une centration sur des critères identiques. De la sorte, les notes commencent à devenir comparables puisqu'elles sont référées à une échelle de valeur identique. Pour parfaire l'homogénéisation des échelles de notation, avant l'enregistrement administratif des notes, chaque correcteur pourrait comparer moyenne et dispersion des notes qu'il a attribuées à l'ensemble des notes dans sa discipline et dans les autres disciplines. Un recalage intra-disciplinaire puis inter-disciplinaire des échelles individuelles de notation, sans nouvelle correction car à classement constant, devient alors possible. D'autres systèmes qui utilisent le barème analytique sont mis en place tel la double correction, ou la correction à plusieurs.

L'ensemble de ces dispositions, dans la mesure où elles sont effectivement oeuvrées, apportent quelque amélioration dans la justesse de la notation. Toutefois, elles ne s'attaquent pas aux conséquences du caractère aléatoire de l'examen ponctuel (cf. chapitre précédent). La proposition ici consiste à vouloir faire considérer le dossier scolaire, ou bien une forme de contrôle continu des connaissances par le jury de l'examen.

L'idée séduit de prime abord: juger le travail effectué pendant une année offre a priori un degré plus élévé de garantie de justesse et de justice que de juger le travail effectué pendant une ou deux journées. La base d'appréciation est relative à l'ensemble du programme, les fluctuations des états psychologiques des élèves se

compensent, il n'y a pas d'effet lié au stress créé par le rituel de l'examen... Pourtant, cela n'est pas évident et à l'analyse les difficultés se pointent en nombre et en gravité.

Tout d'abord, la situation de contrôle continu est moins neutre dans ses formes que la situation de l'examen ponctuel: des épreuves ne sont pas communes à toutes les classes, ni dans leurs finalités, ni dans leurs conditions de passation, ni dans leurs modes de correction, ni dans leurs échelles de notation... Elle est moins neutre également sur le plan psycho-social: si l'examen s'accommode de l'anonymat, le contrôle continu implique la connaissance inter-personnelle de l'évaluateur et de l'évalué. Nous passerons en revue les problèmes que cela pose au chapitre suivant. Rappelons simplement que la représentation que l'on a d'une personne influence directement l'évaluation de ses productions ou de ses capacités et qualités que l'on peut être amené à faire.

A cela s'ajoute le fait que l'évaluation notée, opérée par l'enseignant dans sa classe ou par le formateur dans son stage, subit les effets des mêmes facteurs d'erreurs qui affectent la correction des épreuves d'examens.

En troisième lieu, il faut bien voir que l'enseignant, s'il peut évaluer ses élèves, n'a pas les moyens de les situer par rapport à leurs pairs qui fréquentent les autres classes et les autres établissements. Des enquêtes très larges ont montré une grande variabilité des niveaux scolaires, à classes égales, entre classes et entre établissements: en queue de classe en tel lieu, un même élève se retrouverait en tête de classe en un autre endroit.

En bref, l'examen attentif de cette proposition montre des difficultés pratiques et techniques toutes aussi ardues que celles des évaluations externes. Pour les surmonter, des tentatives d'utilisation de méthodes de modération furent menées en Suisse, en Grande-Bretagne, en Suède. Ces méthodes peuvent s'appuyer sur la passation d'une épreuve commune dans un ensemble de classes du même cours. Les résultats ne sont pas pris en compte d'un point de vue individuel pour classer entre eux tous les élèves des différentes classes. Ils servent à l'enseignant à caler l'échelle de notes qu'il utilise, dans leur moyenne et dans leur dispersion, en situant le niveau de sa classe par rapport aux niveaux des autres classes. Pour cela deux techniques d'ajustement sont disponibles: soit l'attribution des notes respecte certains quotas (dans une classe de niveau supérieur il sera distribué plus de notes supérieures à la moyenne que dans une classe de niveau faible); soit les notes définies par l'enseignant selon ses propres habitudes, avant modération, sont calées par une manoeuvre statistique sur les notes obtenues à l'épreuve commune (on ajoute ou on retranche un nombre de points à toutes les notes). Cette procédure est critiquée selon différents points de vue. L'épreuve commune et les interrogations du maître n'évaluent pas forcément tout à fait la même chose et ne s'appuient pas toujours sur des critères entièrement similaires. Dans ces conditions, des biais sont introduits dans le processus de comparaison. La passation de l'épreuve commune ne se fait en général qu'une seule fois. La répétition de l'épreuve commune coûte cher économiquement, et la réussite à la dite épreuve risquerait de devenir un enjeu focalisant les efforts des différentes classes un peu à la manière des compétitions inter-villes. De la sorte, l'évolution différentielle du niveau des classes n'est pas

intégrée: il n'y a pas de régulation du calage des échelles de notes et à l'arrivée on peut obtenir de nouveau des discordances sensibles.

Dans d'autres circonstances, la modération est amenée au moment de la correction et de la notation d'épreuves académiques internes mais données à l'échelon d'une circonscription par exemple: il s'agit d'un contrôle commun à un ensemble d'établissements. Les copies sont corrigées à deux reprises: l'enseignant corrige et note les devoirs de ses élèves; quelques copies des différentes classes sont revues par un groupe d'experts entrainés à la correction des examens. A partir de leurs résultats un certain réajustement des notes peut être opéré.

Sur le plan global de la métrique, les procédures de modération semblent apporter quelqu'amélioration de la fiabilité des notes: des notes "modérées" ont une valeur prédictive de la réussite scolaire supérieure à celle des notes spontanées. Toutefois la manipulation statistique laisse entier le fond du problème, à savoir les différences de conception de l'évaluation des enseignants et des formateurs, intra et inter disciplines. Et tant que ces différences persisteront, les notes attribuées ne pourront se comparer en toute rigueur. N'est pas abordée, par exemple, la question de l'homogénéisation des objectifs pédagogiques, des contenus effectivement enseignés, des méthodes utilisées, de la matière des évaluations, de leur forme, des critères de correction et de notation...

La réduction des différences d'appréciation peut s'aborder autrement que par la manoeuvre statistique: la proposition est d'organiser une intervention qui puisse limiter les sources de désaccord en égalisant un nombre maxima de facteurs qui organisent la situation d'évaluation et déterminent les critères de correction et de notation. Bien entendu, cette approche sera d'autant plus nécessaire et délicate à mener que l'étendue du champ d'expression offert à l'élève est importante. Concrètement, on a essayé de faire ressortir les principales dimensions sur lesquelles se projetait l'activité évaluatrice. Par exemple, REMONDINO, à partir d'analyses statistiques complexes, isole quatre dimensions principales intervenant dans la correction des compositions françaises: la présentation graphique du devoir, l'usage de la langue vu comme respect de ses règles, le contenu et l'organisation de l'exposé, les aspects personnels tels l'imagination et l'originalité, la maturité. Toutefois, une certaine communauté de dimensions ne renseignent pas sur le poids respectif que chacun des correcteurs leur accorde.

Une autre idée consiste à noter un nombre limité d'aspects sur un nombre limité de parties du devoir. Par exemple, dans le *Test pour le passage en sixième*, l'épreuve de rédaction est notée sur les points suivants: l'ensemble du texte est relatif au sujet, on peut relever au moins trois indicateurs de complexité, il n'y a pas d'erreur de syntaxe ni de segmentation, présence d'une introduction articulée en trois points. Ces choix conservent un aspect arbitraire; la note est référée aux mêmes critères, mais elle ne reflète pas obligatoirement le niveau exact de l'élève dans la mesure où tous les éléments de la compétence ne sont pas abordés.

Une alternative à la fixation d'un nombre limité de critères consiste à multiplier le nombre et le genre des sujets de devoirs proposés aux élèves afin d'élargir la base d'évaluation et de permettre l'expression d'un maximum de qualités

scolaires et individuelles. Il y a ensuite correction par plusieurs évaluateurs afin d'améliorer la fidèlité, la moyenne des appréciations neutralisant en partie les tendances personnelles. Dans les faits, les résultats obtenus par ces deux techniques, l'une plutôt analytique et l'autre plutôt globale, ne se recouvrent pas complètement, peu s'en faut. Elles n'apportent pas de solution entièrement satisfaisante.

Une autre proposition est relative à l'utilisation en pédagogie des techniques d'évaluation utilisées en psychométrie: il s'agit de construire des épreuves standardisées, des tests pédagométriques. Ce dernier point fera l'objet des chapitres neuvième et dixième.

Enfin, la tendance la plus récente de la "docimologie prescriptive" établit explicitement la liaison entre formation et évaluation à travers les taxonomies d'objectifs, la plus célèbre étant élaborée par BLOOM. Nous rejoignons-là le domaine de l'évaluation formative qui sera présenté au chapitre huitième.

Le paradigme docimologique a fait l'objet de nombreuses critiques.

Les travaux de PIERON puis de tous ceux qui ont pu s'en inspirer ne remettent pas en cause l'existence du principe même de la notation ou des examens mais soulignent leurs imperfections, la distance qui les séparent de leurs finalités. Le point d'attaque privilégié est le constat de la variabilité ou de l'imprécision des jugements de valeur. Toutes les conclusions vont dans le même sens: entre des juges reconnus comme compétents surgissent à toute occasion des désaccords sensibles. **La docimologie a démontré de manière forte et incontestable au vu de la ténacité des faits que la notation scolaire était peu valide, peu fidèle et peu fiable.**

Les critiques portent principalement sur les propositions de la docimologie prescriptive. Nous avons vu supra que les mesures envisagées, si elles amenaient des améliorations partielles, n'atteignaient pas le fond du problème de l'évaluation. Comme le souligne avec humour DE LANDSHEERE, les transformations statistiques a posteriori représentent... "des systèmes perfectionnés pour continuer à mal faire les choses".

Pour le dire autrement, les docimologues ont découvert la faiblesse du système de notation puis ils ont travaillé à l'amélioration de ce système sans prendre le temps d'analyser cette faiblesse comme "symptôme" d'un dysfonctionnement de nature psycho-sociale sur lequel la statistique reste sans influence. La docimologie a alerté sur les problèmes de l'évaluation, mais elle aurait échoué dans la recherche de solutions efficaces et équitables. Pour certains auteurs, le paradigme docimologique constitue un sérieux frein aux recherches sur l'évaluation pédagogique dans la mesure où il retarde le moment de poser les vraies questions. Et pourtant, le thème de l'évaluation prend de plus en plus d'ampleur et d'importance au moment où le pouvoir politique annonce que 80% d'une classe d'âge doit parvenir au niveau du baccalauréat. Comment réaliser des progrès sensibles aux niveaux de la pensée et des pratiques pédagogiques lorsqu'on connaît le poids considérable des habitudes mentales et sociales ?

Une trop grande diversité des approches engendrerait un risque de confusion, une difficulté à isoler les éléments les plus pertinents... Une technicisation excessive dirigée sur des procédures statistiques sophistiquées écarteraient les formateurs et les enseignants de la recherche des solutions.

En ce qui concerne la docimologie, il convient de ne pas lui faire dire plus loin que sa compétence: élargir son champ d'application de l'examen à l'évaluation en général fait courir le risque d'un réductionnisme malgré tout bien commode qui situe la raison des processus sociaux en jeu dans l'évaluation au niveau du simple traitement des procédures! Dans ces conditions pourquoi se casser la tête à vouloir, par exemple, améliorer la qualité de son enseignement ou moduler sa relation éducative si la solution se trouve tout bêtement dans le nombre de corrections à faire d'une même copie, ou dans le choix d'une système de modération des notes... La logique du réducationnisme docimologique, poussée à son extrêmité, conduit à confier toute l'évaluation, dans l'ensemble de ses dimensions à un corps de spécialistes de la mesure exacte.

Sur un plan idéologique, la docimologie en cherchant à améliorer les examens les consolide, en accord avec les lois du fonctionnement économique et de la stratification sociale, en renforçant l'idée de méritocratie... L'hypertrophie de la durée et de la fréquence des examens peut amener des perversions telles le "bachotage" qui focalise les investissements sur la note au détriment des apprentissages véritables. De même, nous verrons au chapitre suivant comment l'omniprésence d'un système de contrôle est l'une des raisons qui bloquent les évolutions pédagogiques.

DOMINICE écrit, et ce sera la conclusion de ce chapitre :
"La docimologie a fourni des techniques mais elle n'a pas réussi à modifier la fonction politique et sociale de l'évaluation. La rationalisation à laquelle elle est parvenue peut ainsi être considérée comme technocratique, c'est-à-dire qu'elle déplace sur des spécialistes de la mesure le pouvoir de discrimination jusqu'alors exercé par chaque pédagogue à l'intérieur de sa classe.
De fait, les pistes de recherche les mieux adaptées à la démocratisation scolaire des lycées ne sont pas celles qui conduisent à des évaluations pléthoriques, mais plutôt celles qui redonnent à l'évaluation sa véritable place comme accompagnatrice et comme moyen de régulation de l'action pédagogique".

Cette pensée nous amène à nous interroger sur la valeur formatrice de l'évaluation.

CHAPITRE 6

L'EVALUATION EST-ELLE COMPATIBLE AVEC LA PSYCHOPEDAGOGIE ?

L'évaluation, que ce soit dans ses pratiques ou dans ses principes de base, a été l'objet de nombreuses critiques. Dans les deux chapitres précédents, nous avons abordé les questions soulevées par l'approche docimologique de l'évaluation, approche essentiellement centrée sur les aspects métrologiques. Un autre ordre de critiques est relatif aux relations qui s'établissent entre les pratiques de l'enseignement ou de la formation, les pratiques de l'évaluation, et leurs effets de nature psychologique sur les élèves ou les stagiaires. La focalisation sur le jeu de ces relations amène, sur un plan plus général, à poser la question de la compatibilité des finalités pédagogiques des systèmes de formation et des effets psychologiques des systèmes d'évaluation. La formulation de cette question dans le sens d'une problématique invite à regarder quels sont les effets psychologiques de l'évaluation sur le formé, d'apprécier la valeur positive ou négative de ces effets par rapport à l'atteinte des objectifs pédagogiques poursuivis par le maître ou le formateur: l'évaluation apporte-t-elle une aide aux efforts pédagogiques de l'enseignant en stimulant psychologiquement l'élève, ou bien au contraire l'évaluation contrecarre-t-elle les efforts magistraux par des effets de dégradation psychologique sur l'apprenant ?

Nous allons essayer d'enregistrer et d'analyser les arguments avancés pour justifier les deux positions extrêmes tout en prenant soin de ne pas perdre de vue que le choix d'un objet, d'une procédure, d'une méthode, d'un outil d'évaluation n'est jamais neutre: il s'appuie toujours sur une certaine conception idéologique, sur une certaine philosophie de l'éducation. A travers les argumentations avancées, il est important de rechercher la trame idéologique de leur genèse, de leur contexte d'élaboration, tant il est vrai que c'est le contexte qui donne le sens.

Considérons en premier les points de vue qui postulent que l'évaluation scolaire représente un stimulant appréciable des dynamismes intellectuels et motivationnels qui sous-tendent les apprentissages.

Les pédagogies dites de la motivation posent comme axiome l'existence d'une relation directe entre évaluation et motivation. Le seul fait de prendre connaissance de la valeur de sa production satisfait le besoin de jugement et de reconnaissance sociaux de la personne. Un jugement favorable renforcerait le sentiment de compétence et par contrecoup la motivation à apprendre; un jugement défavorable constituerait un défi à relever, défi libérateur des ressources internes nécessaires. Il est certain que l'exercice de la seule auto-évaluation, c'est-à-dire le recours unique à un système d'évaluation auto-référencé, ne fournirait pas à l'apprenant les données suffisantes pour une saine régulation de ses apprentissages. (Nous entendons par une saine régulation celle dont le produit prend forme d'un système de normes intériorisées susceptible de générer la souplesse adaptative des comportements tout en respectant les aspirations relatives à la "vraie nature" de l'individu). Comme l'énonce NUTTIN, "... le résultat éprouvé par le sujet comme terme de son comportement est avant tout le résultat tel qu'un autre l'apprécie et lui communique".

A partir de là, on peut postuler l'existence d'un besoin profond de confrontation de son oeuvre à l'opinion d'autrui et ceci en réponse à l'exercice de trois fonctions au moins. Tout d'abord, l'évaluation amène une facilitation et une accélération des apprentissages: pour BARBIER, plus un individu est évalué positivement et plus il produit la bonne réponse avec facilité et rapidité, c'est-à-dire qu'il intègre l'objectif comportemental correspondant. L'évalué fait sans doute une démarche de réappropriation du contenu du jugement de valeur en tant qu'informateur de ce qu'il faut faire ou ne pas faire pour réussir une tâche. Le renforcement positif rend l'élève disponible et ouvert pour l'intériorisation et la cristallisation du schème de réponse, le renforcement négatif empêchant la stabilisation des comportements inadaptés.

La seconde fonction est celle relative à la prise de rang social. Il est trivial de constater que toute activité humaine, et même les activités de détente, de loisirs ou sportives, donne lieu à classement, à catégorisation hiérarchique et que les individus recherchent ces classements: il suffit de penser à toutes les manifestations, à caractère non professionnel et non obligatoire, qui rassemblent joueurs de pétanque, coureurs à pied, danseurs, possesseurs d'animaux ou de voitures anciennes, joueurs de bridge ou du scrabble... pour joutes, concours et tournois aux prix bien souvent dérisoires eu égard à la somme des efforts déployés pour acquérir et tenir son rang. La recherche active de son rang sur l'échelle des performances (et donc sur l'échelle sociale) constitue un puissant levier de dynamisation de l'activité et des apprentissages, levier qui prend appui sur une émulation tant individuelle que collective.

Enfin, la troisième fonction psychologique à laquelle répond le besoin d'évaluation est une fonction d'orientation: pour REUCHLIN, l'évaluation éclaire l'élève sur son devenir, elle est nécessaire à la régulation de la scolarité et des cursus, à la prise de décision relative aux choix de carrière professionnelle.

Bien évidemment, des opinions radicalement hostiles à l'évaluation et étayées par une argumentation relative à la psychopédagogie sont en circulation sur le marché des idées.

Un premier objet de critique touche aux motivations que stimulent les pratiques de l'évaluation traditionnelle: ces motivations n'auraient pas forcément les vertus escomptées par le pédagogue mais présenteraient un aspect plutôt négatif. Par exemple, trop souvent, l'évaluation traditionnelle reposerait sur la peur de l'échec plutôt que sur le désir de réussite. Cette peur de l'échec irait croissant avec l'importance de l'enjeu social et serait maximale au moment des examens, des concours déterminants pour la définition de la trajectoire d'insertion sociale. Elle engendre un stress négatif dont les effets sur l'économie psychosomatique – anxiété, nervosisme, anorexie ou boulimie, perte du sommeil, perte des cheveux, ulcérations, eczémas... – peuvent se révèler gênants, voire désastreux. A la limite, dans un pays comme le Japon, l'échec d'une vie scolaire trouvera son issue dans le suicide.

Les conséquences psychologiques de l'exaspération des investissements dans la compétition scolaire sont également dénoncés comme une perversion du système d'évaluation. Essentiellement, la compétition scolaire met l'individualisme au premier plan des valeurs transmises au détriment de l'esprit d'entraide et de solidarité. Elle peut amener des comportements uniquement égocentriques, fermer les communications horizontales entre pairs, rendre inapte au travail en équipe que pourra demander ultérieurement l'entreprise ou la vie associative et sportive. Probablement aussi grave, la compétition excessive liée à la sélection par le savoir scolaire aurait comme effet de retournement une déculturation progressive des personnes en formation. Prenons l'exemple de l'enseignement des mathématiques qui a supplanté le latin comme matière de sélection – orientation.

JACQUARD analyse ainsi la situation et ses conséquences: perçues et enseignées dans la seule direction de leur évaluation formelle, les mathématiques fonctionnent comme un ensemble de techniques mécaniques et désincarnées, sans mise en perspective, sans transfert de la théorie à l'expérience, donc sans culture; "Mais l'inculture qui baigne les mathématiques a aussi des conséquences hors du champ scolaire. Puisqu'on privilégie le lettrisme et l'efficacité mécanique en maths, plus personne ne s'interroge sur le sens ni sur la pertinence de tel sondage, de telle statistique, de telle mesure". Nous ajouterons que l'absence de formation à l'attitude réflexive, à la projection et à l'analyse des significations culturelles, philosophiques et esthétiques, au sens critique et à la critique du sens conduit tout droit au retour d'un rapport magique au savoir qui ouvre le chemin de la crédulité.

Plus radicalement encore, on a pu conclure que la focalisation excessive sur l'évaluation opère des distorsions importantes dans la formation de la personnalité de l'apprenant, quelque soit son âge. L'évaluation ne permet d'apprécier qu'une fraction des talents de l'individu. La part des talents non pris en compte, si elle n'est pas développée dans un autre contexte institutionnel, va progressivement s'éteindre et conduire à un appauvrissement dommageable. Par exemple, dans les cours de l'école primaire, la connaissance d'une langue étrangère, ou d'un langage informatique, ou d'un instrument de musique... n'est pas reconnue. D'une manière générale, les institutions de formation ne prennent jamais en compte d'une manière

systématique, explicite et officielle ce que l'on appelle les "acquis expérientiels". Dans le même ordre d'idée, la non prise en considération des activités et des acquisitions spontanées, en milieu naturel, combinée à la centration des intérêts, des représentations, de l'attention sur les seules valeurs scolaires reconnues par l'évaluation écarte par contrecoup indirect, ou parfois par des pressions explicites, le plaisir d'apprendre dans des situations au caractère exploratoire, dans des situations de production faisant appel à l'initiative personnelle... La frustration qui en découle peut s'accompagner de véritables mutilations psychologiques dans le domaine de la construction de l'identité psycho-culturelle: rappelons-nous qu'à l'école coloniable en Martinique le parler créole se sanctionnait par le port du bonnet d'âne; plus près de nous, le rejet de leur culture d'origine par l'école participe aux difficultés existentielles des jeunes immigrés... Les mutilations psychologiques se réfèrent également au plan des capacités cognitives. Par exemple, l'école et autres organismes de formation pour adolescents et adultes consacrent la majorité du temps d'étude à proposer l'acquisition de démarches de résolutions convergentes qui procèdent par réduction progressive des données jusqu'à l'obtention d'un résultat unique. Les démarches divergentes, qui font rechercher un ensemble de possibles, et qui sont à la base de l'imagination poétique ou fantaisiste, des explorations scientifiques, de la créativité, de l'invention, et plus simplement de la débrouillardise ne sont pas enseignées; leurs correspondants en termes de structures et de schèmes mentaux ne seront donc pas mis en place.

Un dernier ordre de critiques des pratiques d'évaluation touche au renforcement de l'idéologie dominante qui rend légitime les différences interindividuelles. Cette idéologie s'énonce synthétiquement dans la formule de l'égalité des chances et aussi dans son complément logique, l'égalité des malchances. Elle conjugue le droit de nature -l'idéologie des dons- et le droit de culture -la méritocratie-. Cette idéologie renvoie donc l'explication du degré de réussite -ou d'échec- en première lecture à l'hypothétique existence d'un don et en seconde lecture à la manière dont ce don est travaillé qualitativement et quantitativement. Elle est fort commode lorsqu'on l'applique pour "expliquer" la réussite d'un tiers dans la mesure de l'aspect sécurisant de sa redondance: dire de quelqu'un qu'il a réussi car il est le meilleur, le plus méritant des mieux doués naturellement satisfait tous les besoins de rationnalité.

Mais comment rendre compte de son propre échec en appliquant cette idéologie à soi-même? (Et comment ne pas l'appliquer puisqu'elle domine les esprits?) Si je n'ai pas réussi, c'est parce que je ne suis pas doué et/ou parce que je n'ai pas suffisamment travaillé. Dans les deux cas, je suis responsable de mon échec. Et l'on peut voir alors les élèves en difficulté retourner le jugement externe sur eux-mêmes, développer un sentiment d'incompétence, un sentiment d'indignité et de culpabilité, un sentiment d'étrangeté par rapport à la culture et aux normes scolaires qui éventuellement se généralisent comme sentiment antisocial. Il en résultera parfois une inhibition de l'investissement dans l'activité scolaire et plus sûrement une inhibition concomminante du développement des fonctions cognitives.

Une autre manière de poser le problème des relations entre l'évaluation et la psychopédagogie consiste à souligner leurs ambiguïtés, leurs contradictions, leurs paradoxes.

Ainsi, on constate que les élèves ou les stagiaires sont demandeurs d'évaluation, que cette demande se révèle la plus forte dans les systèmes les plus sélectifs, qu'elle se manifeste même dans les cas où l'élève est quasiment sûr du résultat qu'il va obtenir. On peut penser que l'évaluation assurerait une sorte de fonction d'assurance, ou de processus itératif de réassurance de l'identité vue dans ses aspects de classement social, de performance et de compétence, de l'image de soi. On serait finalement proche d'un phénomène de recherche de consonnance cognitive relativement au sentiment que l'on s'est construit de sa compétence... ou de son incompétence.

Ceci amène à poser le problème de l'émergence et du développement des motivations à se changer alors qu'il semblerait que le cycle évaluation – renforcement de l'identité individuelle aboutisse à maintenir la personne dans un même état de dispositions psychologiques: pour le bon élève qui a, au fond, peu de raisons objectives de vouloir se changer cette forme de conservatisme n'a pas valeur de handicap; pour l'élève aux résultats faibles, ce qui devient une cristallisation de la résistance au changement au fur et à mesure des résultats des évaluations successives va laisser des traces durables à l'âge adulte. Par exemple, une étude de ROUSSET montre que les personnes qui sont le moins satisfaites de leur poste de travail sont celles qui désirent le moins se former et réciproquement: en fait, on peut dire que la motivation à se former est liée au sentiment de satisfaction et de sécurité qui permet d'affronter la perspective du changement.

Une autre étude de CUSIN et PIOLAT conclut que l'existence d'un passé scolaire ressenti comme défectueux entraîne une sous – évaluation systématique des performances chez les sujets concernés. Si blocage et inhibition sont des incapacités fonctionnelles de progresser dans certains domaines, on peut dire que le vécu individuel négatif des procédures d'évaluation est un facteur efficace de blocage et d'inhibition des progrès cognitifs chez l'enfant et chez l'adulte, alors que ces derniers restent demandeurs de telles procédures.

Il n'existe pas seulement que des procédures d'évaluation formelles et quantifiées. Il y a également à prendre en considération l'ensemble des jugements de valeurs énoncés par le maître ou le formateur et qui imprègnent de manière continue la relation éducative et la communication pédagogique. En considérant le but officiel de l'éducation et de la formation qui est la transformation positive et utile socialement de la personne, on devrait s'attendre à n'observer que des jugements de valeurs qui puissent rendre un service à l'apprenant. Pourtant, la simple observation d'un bulletin de notes trimestriel, ou la simple lecture des commentaires portés sur la feuille de devoir montre que l'évaluation scolaire distribue des jugements négatifs en accompagnement des notations chiffrées. Lorsque ces jugements négatifs se répètent tout du long de la scolarité et se prolongent au cours de la formation professionnelle, alors les effets sont bien souvent désastreux en ce qui concerne la formation du sentiment de compétence.

De plus, dans la relation éducative qui fonctionne au jour le jour, minute après minute, le maître et le formateur projettent, parfois volontairement et le plus souvent inconsciemment, par les canaux de la communication verbale et de la métacommunication (ton de la voix, émotion partagée, posture, mimique...) un jugement

de l'élève ou du formé. Si les jugements positifs, surtout lorsqu'ils se répètent, amènent des sentiments de sécurité, de confiance en soi, d'affirmation de soi,... les jugements négatifs amènent sentiments d'échec, d'humilation, d'infériorité: la dévalorisation externe s'intériorise en forme d'auto-dévalorisation. Ce processus est d'autant plus accentué que le jugement glisse de l'oeuvre à la personne. On n'évalue plus le nombre et la nature des erreurs d'orthographe, mais les qualités individuelles, psychologiques et comportementales de l'auteur. Ce n'est plus seulement le niveau mathématique qui est faible, mais aussi les capacités intellectuelles du sujet.

D'une manière paradoxale, on a pu observer que ce sont les interactions positives qui s'adressent le plus fréquemment aux bons élèves, c'est-à-dire à ceux qui a priori en ont le moins besoin. POSTIC propose l'interprétation suivante de ce phénomène :
"Si on analyse en profondeur la nature des interventions de l'enseignant, on relève des différences fondamentales: à ceux qui sont perçus comme "bons élèves" reviennent les actes pédagogiques ayant une finalité de stimulation, d'encouragement, de support, à ceux qui sont perçus comme "mauvais élèves", les actes destinés à faire accélérer le rythme, à critiquer, à assurer la discipline. Les contacts conflictuels avec ces derniers provoquent des réactions de domination de l'enseignant, qui cherche à assurer une compensation en prodiguant aux autres des manisfestations d'intérêts, de sympathie. L'enseignant utilise le jeu compensatoire pour établir un certain équilibre du groupe, pour se rassurer lui-même et se confirmer dans l'idée de l'efficacité de son action pédagogique".

De ce point de vue du fonctionnement ordinaire de la relation éducative, on peut donc dire que l'évaluation participe à ce que j'ai appelé ailleurs la double contradiction et le rejet mutuel: contradiction vécue par le formateur entre sa mission éducative et sa conception d'élèves diminués, défavorisés, handicapés, perçus comme étant incapables d'apprendre; contradiction vécue par le formé dans un sentiment qui mèle respect, attirance, désir et méfiance, peur de l'institution de formation. Ces processus vont être renforcés par la réalisation automatique des attentes, autrement désignée sous le nom d'"effet Pygmalion": une fois que nous n'attendons plus rien de bon d'une personne, nous allons aligner nos actes sur cette croyance.

L'ensemble de ces contradictions est probablement réductible à une contradiction générique que LURCAT perçoit et dénonce...
"L'ambiguïté fondamentale d'un système compétitif qui est amené à sélectionner sur ce qu'il doit transmettre, en créant des obstacles qui s'ajoutent à la difficulté propre aux connaissances à transmettre".

Comment lever cette contradiction fondamentale qui est au principe du mélange des genres ?

Il faut en premier lieu admettre la nécessité de l'évaluation; en effet comment servir correctement une fonction pour laquelle on n'aurait que peu d'estime, sinon de l'opposition? Encore conviendrait-il de mettre l'évaluation à sa juste place, c'est-à-dire au service des objectifs éducatifs et de la psychopédagogie, et donc en dernier ressort au service des évalués. Pour ce faire, un certain nombre de

précautions doivent être prises et un certain nombre de pièges évités. Par exemple, l'évaluation sommative doit être externe, c'est-à-dire proposée et mise en oeuvre par un jury extérieur à la classe et à ses maîtres dans la mesure où on ne peut pas être juge et partie. A l'interne le développement de l'évaluation formative en favorisant le centrage sur le processus d'élaboration des réponses plutôt que sur leur justesse considérée comme un absolu devrait favoriser la professionnalisation des enseignants comme éducateurs, comme formateurs plutôt que comme censeurs. Le recours à une auto-évaluation guidée et co-opérative, qui apprend à choisir et à utiliser les critères d'appréciation, peut déboucher sur la fixation d'objectifs pédagogiques qui soient partagés. De la sorte, enseignant et enseigné apprennent à se connaître et de l'évaluation du travail ils passeront plus ou moins rapidement à l'estime réciproque. L'évaluation commence, à partir de cet instant, à devenir positivement signifiante, et elle le sera d'autant plus que le formateur va modifier son mode d'intervention en fonction des résultats que l'évaluation met en évidence. C'est à partir du moment que l'enseignant aura fait comprendre à l'élève son intention profonde et sincère de participer à son développement cognitif et à l'épanouissement de ses talents et que l'évaluation n'est que l'un des moyens de concrétisation de cet idéal que l'on pourra dire que l'évaluation est compatible avec la psychopédagogie. D'une manière plus large, nous pensons que les enseignants et les formateurs, sur le plan de leur perfectionnement technique dans la manipulation de la relation éducative, tireront de grands bénéfices à se former aux techniques de la "médiation" telles que les présentent FEUERSTEIN par exemple. On peut également supposer que lorsque les idées et les résultats des expériences véhiculés par les travaux relatifs à l'éducabilité atteindront un niveau de généralisation suffisant les représentations des formateurs évolueront dans le sens d'une conception plus ouverte quant aux possibilités de transformation des aptitudes de leurs élèves. Cette question mérite à elle seule l'élaboration d'un volume.

CHAPITRE 7

L'EVALUATION PEDAGOGIQUE MASQUE-T-ELLE UN TRI SOCIAL ?

Nous avons successivement considéré la critique des pratiques d'évaluation du point de vue métrologique et du point de vue des conséquences psychopédagogiques de leurs effets. Le troisième ordre de critique, après la mesure et la psychologie, touche aux conséquences sociales de l'évaluation.

Il est facile de constater que l'évaluation pédagogique intervient de manière très active dans les processus d'orientation et de sélection. En réaction à ce fait, deux thèses opposées s'affrontent. Pour les uns, le tri social par l'évaluation est chose nécessaire; pour les autres, l'évaluation pédagogique ne serait qu'un instrument de sélection aux mains des classes dominantes.

Les tenants de la première thèse utilisent trois types d'arguments: selon REUCHLIN la fonction de tri social de l'évaluation pédagogique procède de trois ordres d'impératifs. En premier lieu la hiérarchisation des qualifications professionnelles impose la différenciation des individus d'après leur niveau général de formation de base. Si le système d'enseignement et de formation ne se préoccupait pas de cette différenciation, d'autres s'en chargeraient forcément et à l'aide d'échelles de valeurs plus ou moins satisfaisantes. De fait le diplôme défend la valeur sociale de la formation acquise; le supprimer ne serait-il pas préjudiciable à ceux qui ne disposent pas d'autres moyens pour s'affirmer sur le marché de l'emploi?

En second lieu, la préoccupation de garantir la sécurité des individus ou des collectivités nécessite l'organisation de l'accès aux emplois, et ce d'autant plus qu'ils comportent des responsabilités importantes. On ne peut pas laisser n'importe qui conduire une machine ou un véhicule de forte puissance, prescrire des médicaments aux principes très actifs, diffuser des nouvelles à l'échelon national en très peu

de temps... L'institution qui a pris en charge la formation d'un professionnel doit s'assurer de sa compétence avant de lui permettre l'accès à l'emploi.

En troisième lieu, il est nécessaire de contrôler la rentabilité des ressources investies dans l'appareil éducatif, ressources qui absorbent une large part du budget national. La bonne gestion des fonds publics consacrés à la formation demande d'offrir, à chaque niveau, des enseignements à ceux qui sont en mesure d'en tirer profit parce qu'ils en maîtrisent suffissamment les pré-requis. Il y a alors obligation de contrôle pédagogique dans le cadre de l'orientation-sélection.

Les spécialistes de l'analyse institutionnelle rejettent ce type d'arguments. Leur rejet peut s'exprimer parfois de manière radicale. Par exemple LOBROT n'hésiste pas à affirmer qu' "... il y a incompatibilité entre la formation et la pratique sociale du contrôle-sanction... (l'Etat) infuse (à la collectivité) une conception des rapports sociaux fondés sur la méfiance et le contrôle policier". Le rejet peut faire place au questionnement et GUIGOU tente de cerner l'enjeu institutionnel de l'examen :
"... l'évaluation permet-elle l'appropriation collective et intellectuelle du savoir, favorise-t-elle la création d'un nouveau savoir social et donc l'accession des formés à des nouveaux rapports pédagogiques et sociaux? Ou bien cherche-t-elle à contrôler la diffusion du savoir institué pour mieux maîtriser la sélection, renforcer implicitement la séparation entre savoir et non-savoir, garantie essentielle de conservation des institutions éducatives et donc des institutions politiques?"... et en dernier lieu, ajoute le sociologue PASSERON, de l'ordre social :
"La fonction de sélection et de solennisation de la sélection au travers de l'examen n'est plus, dès aujourd'hui, immédiatement déductible des besoins techniques de l'économie et on ne comprendrait pas que des sociétés comme les nôtres attachent autant de conséquences aux examens si on ne mettait pas en rapport les examens avec les besoins sociaux des classes privilégiées pour qui l'effet de certification par l'examen des privilèges sociaux et culturels qu'elles possédaient dès le départ contribue à la perpétuation et à la légitimation de leur position dominante".

Sur le plan de la relation éducative, MEIGNIEZ et VALATX proposent l'analyse suivante des finalités de la notation :
"Bien souvent, la notation n'est pas là pour apprécier des connaissances ou des aptitudes, mais plutôt pour signifier la domination de l'idéologie du professeur sur celle de l'élève qui se voit contraint d'étouffer la sienne: la notation devient un instrument pour démontrer son pouvoir, et non pour mesurer les performances de l'élève".

En bref, toute procédure d'évaluation viserait à légitimer d'une part la sélection sociale, voire la reproduction des classes sociales et d'autre part les rapports de subordination entre formateurs et formés.

On perçoit aisément la distance qui sépare ces deux points de vue du même objet évaluation. Il n'est jamais confortable de s'interposer entre des polémistes extrémistes. Pour nous forger une opinion personnelle, il faut une fois de plus tenter de se rapprocher de la réalité pour mieux comprendre comment fonctionne le tri social à travers les pratiques de l'évaluation.

Pour BARBIER, il existe de multiples fonctions déclarées de l'évaluation des agents. Cette multiplicité de fonctions déclarées, parfois contradictoires, sert principalement à deux choses. Premièrement, elle permet de légitimer les différents contrôles évaluatifs aux yeux de ceux qui en sont l'objet.

Secondement, elle occulterait la fonction réelle de ces contrôles. La question devient alors: comment accéder à cette fonction réelle des pratiques de l'évaluation? Pour y répondre, l'interrogation va se centrer sur la compréhension des contextes sociaux de la mise en place et du maintien des pratiques d'évaluation.

Le premier de ces contextes est celui de la réglementation d'un marché des biens et des services. C'est un contexte qui est historiquement premier. Que l'on pense, par exemple, au système des corporations, où plus près de nous aux procédures d'accès à certaines professions libérales, aux règlementations de la mise en place des supermarchés, aux agréments accordés aux professionnels par les compagnies d'assurances... Dans ce contexte, les procédures d'évaluation se prolongent plus particulièrement par l'attribution de divers habilitations, autorisations, homologations, agréments, qui ouvrent la possibilité de l'exercice de l'activité. Il y a par ce biais contrôle de l'accès au marché, ce qui lui confère en contre-partie une certaine protection.

Le second contexte est relatif au fonctionnement des systèmes d'activité, tels les entreprises industrielles et/ou commerciales par exemple. Ces systèmes d'activité présentent trois caractères à prendre en compte pour saisir leurs liens avec la question du tri social par l'évaluation. Ils supposent une division du travail qui se traduit dans une spécialisation des tâches. Ces tâches sont organisées d'une manière fonctionnelle et hiérarchisée de telle sorte qu'elles puissent aboutir à un même résultat d'ensemble. Ces systèmes instaurent une distinction entre le contenu des tâches et les individus qui les exercent (séparation de la fonction et du statut). On retrouve cette différenciation qui à la fois amène et justifie la sélection tout aussi bien dans le champ de l'exercice du travail (hiérarchie des qualifications à alimenter par exemple), que dans le champ de la formation (niveaux et filières d'enseignement dans lesquels il faut orienter les élèves) ou que dans le champ de la production des moyens d'existence (face à la hiérarchie des qualifications est disposée une grille de salaire).

Par rapport à l'alimentation en personnes des strates et catégories de ces systèmes d'activité, les procédures d'évaluation manifestent toute leur importance au niveau des itinéraires individuels. Ainsi, le diplôme participe à l'attribution du statut social; dès l'entrée dans la vie active notes, examens sanctionnent la formation continuée et participent dans une certaine mesure à l'attribution des promotions.

Ceci est tellement vrai que l'on peut remarquer au passage que les activités de formation se divisent en deux catégories: les actions qualifiantes qui déterminent la situation sociale et professionnelle ou son changement; les actions de développement qui perfectionnent les compétences individuelles à l'intérieur d'une catégorie sociale, sans perspective d'en changer. Dans le cas de la première catégorie, on observe une augmentation de la fréquence et une diversification des moments de

contrôle, principalement dans les classes ou les stages préparant des diplômes susceptibles de jouer en termes de niveaux d'insertion dans l'emploi. Dans le cas de la seconde catégorie, on ne constate que peu ou pas du tout de procédures d'évaluation, comme par exemple dans les classes de fin de scolarité obligatoire, ou dans les "stages parking" de traitement social du chômage, ou encore dans les perfectionnements professionnels qui n'ont d'autres finalités que l'épuisement des fonds du 1%, ... autant de formules sans enjeux sociaux du fait de l'absence de système de certification reconnue.

Toujours dans ce second contexte, une autre remarque attire notre attention, remarque qui renforce l'idée que la fonction première de l'évaluation pédagogique serait bien la distribution des individus sur l'échelle des catégories socio-professionnelles, et sur la hiérarchie des statuts dans une même catégorie: le statut social est fixé dès l'entrée dans la filière et non à sa sortie.

L'exemple le plus frappant reste bien celui de la préparation des concours d'entrée dans les grandes écoles. Le plus problématique n'est pas d'obtenir le diplôme d'ingénieur, mais de rentrer dans le cursus de formation. Ainsi, les étudiants en médecine sont sélectionnés sur la base de leurs connaissances académiques sans aucune prise en compte de leurs qualités relationnelles... En généralisant de telles observations, on peut en conclure avec BARBIER que :
"... pour pouvoir exercer bon nombre d'activités sociales, il faut déjà se trouver dans le champ social d'exercice légitime de ces activités". Notons ici que l'appartenance à ce champ se décide certes au vu de connaissances dont la maîtrise peut être objectivée, mais aussi au travers de la perception de ce que BOURDIEU et PASSERON appellent le système des impressions convergentes portant sur le système de manières caractéristique d'une position sociale. Joue également dans le jeu de la reconnaissance et de la légitimation tout le poids des structures dominantes, principalement sociale -importance du groupe social d'appartenance: à niveau scolaire égal les enfants des milieux ouvriers bénéficient de moins de pronostics favorables- et sexuelle -à niveau de qualification égale, dans l'entreprise les femmes auront moins souvent un poste à responsabilité et seront plutôt moins payées-.

Au niveau technique, les pratiques d'évaluation en usage dans ce contexte prennent souvent en compte des éléments analytiques tels que les notes scolaires, les résultats obtenus à des tests, les appréciations qualitatives, des rapports de stages... Ces éléments s'intègrent dans un dossier récapitulatif, dossier de formation, livret scolaire, diverses fiches de suivi..., dossier où peut s'inscrire également les données du curriculum vitae ou du bilan professionnel. En quelque sorte, ces pratiques se focalisent sur le profil des capacités des évalués. Le produit de l'évaluation s'exprime en forme de prévisions de réussite ou d'itinéraire, de programmes de développement personnel, de projets de progression ou d'appréciation d'un potentiel, l'objectif de transformation personnelle étant identifiable en terme de niveau à atteindre.

Les processus psychologiques rejoignent alors les effets de tri social de l'évaluation. Par exemple, une série de mauvaises notes accompagnées d'autant de jugements de valeur négatifs ferment toute perspective en dégradant le sentiment de la

compétence, en s'attaquant à l'image de soi, en réduisant la probabilité subjective d'accès à tels ou tels fonction ou statut.

Bien entendu, il faut rappeler ici que l'intériorisation des jugements de valeur négatifs joue au moment des évaluations explicites et organisées, soit dès la passation des épreuves quand le sujet sent, s'aperçoit qu'il "sèche", soit au moment des rituels de restitution des résultats, et aussi à chacune des interactions entre le maître et l'élève, entre l'agent de maîtrise et l'ouvrier, entre l'employé et le chef de bureau... La dévalorisation de l'autre qui commande la mise à sa place va utiliser tous les canaux de la communication et de la méta-communication (ton, mimique, regard, attitude...). On retrouve là le plan d'intersection entre l'évaluation dans la psychopédagogie (abordé au chapitre précédent) et l'orientation socio-professionnelle. A ce propos on parlera de sélection par l'échec plutôt que de sélection par la motivation et libre choix.

Le dernier contexte dégagé par BARBIER est celui de la gestion socialisée d'un intinéraire de formation qui recouvre un ensemble de pratiques plus récentes. Le dispositif d'accueil, d'information, de suivi, d'orientation est bien installé dans le système scolaire (palliers d'orientation institutionnellement établis, corps de conseillers professionnel...). Au niveau des adultes, on peut évoquer le rôle des psychologues de l'A.F.P.A. au moment de l'examen psychométrique des candidats aux formations professionnelles, à l'occasion du suivi des sections dites préparatoires et des formations organisées en modules. Enfin, les entreprises les plus en pointe dans la gestion de leur ressources humaines n'hésitent plus à établir un dispositif de gestion prévisionnelle du personnel alimenté par des pratiques d'évaluation.

Quelque soit le contexte, apparait donc comme régulier le fait que l'évaluation pédagogique formelle et ritualisée de la personne intervient à chacun des moments de la vie où se décide et éventuellement s'infléchit la trajectoire de l'insertion socio-professionnelle. Ceci confirme l'évidence de la fonction de tri social des pratiques de l'évaluation institutionnelle. Il reste maintenant à essayer de comprendre le processus fonctionnel en jeu au moment des examens mais aussi, et peut-être principalement, dans la relation socio-éducative qui opère au jour le jour.

Nous avons évoqué plus haut le constat de la sélection par l'échec, par l'élimination plutôt que par le libre choix. Curieusement, cette manière de procéder ne soulève guère de protestations de la part des usagers du système scolaire ou parmi les demandeurs d'emploi, ou parmi les salariés non-promus. Cette acceptation concensuelle pourrait indiquer que l'ensemble du corps social a intériorisé la hiérarchie des valeurs reconnues et validées par l'école et l'entreprise en tant que système de valeurs universel dont le degré de possession (et de distanciation) conditionnerait le degré de réussite individuelle. On peut supposer deux voies principales à cette intériorisation. La première est corrélative au procès d'imprégnation qui passent par les canaux de communication usuels comme les médias habituels, les discusions "naturelles" entre pairs ou dans le milieu familial, la lecture des journaux et livres non-spécialisés... Dans ce cas, l'individu évolue dans un "bain de valeurs" dont il absorbe les principes actifs sans aucune conscience de ses propres changements qui peuvent en résulter. La seconde situation, à l'école et au travail, ajoute au bain naturel une volonté pédagogique, ou hiérarchique, assurée

par un enseignant, un formateur ou un contremaître, d'induire chez l'individu les transformations attendues car conformes au modèle, au stéréotype social, aux systèmes de valeurs et de normes majoritairement admises.

La mise en forme psychologique dans son rapport à l'appartenance et à la trajectoire sociales combine de façon interactive l'incorporation des valeurs structurées par l'inconscient collectif et véhiculées dans la relation sociale commune au travail d'inculcation des principes pédagogiques et techniques transmis par l'enseignant, le formateur, puis ensuite par la hiérarchie de proximité. Les moments d'évaluation formelle et institutionnelle vont permettre de contrôler et de valider la conformité du "produit psycho-social" que représente l'élève et le travailleur au standard du groupe auquel il souhaite appartenir. Cela sert à dire qu'
"... évaluer c'est donc exercer un pouvoir sur d'autres afin de les faire agir conformément au système de normes de la société considérée". (BEAUVOIS & ENRIQUEZ).

On pourrait d'ailleurs longuement discuter sur les conséquences de la déconnection des valeurs scolaires de ce qui assure leur génération à savoir l'utilité concrète du savoir: le besoin d'apprendre, de lire par exemple -voir les enquêtes menées par le Ministère des Armées sur l'illettrisme- va disparaître avec le besoin d'évaluation une fois que les "jeux sont faits", c'est-à-dire une fois que l'insertion sociale est effective.

Curieusement, on retrouve une déconnection de même nature dans les entreprises majeures et dans les administrations, déconnection produite par les processus de bureaucratisation. L'acte administratif devient alors uniquement référencé à la hiérarchisation des valeurs spécifiques de l'institution et il va en résulter une perte de sens social de l'activité. Tout un chacun s'est au moins une fois au cours de son existence heurté au non-sens bureaucratique.

Une étude de GILLY a permis d'avancer dans la compréhension de la structure du système de représentations à l'oeuvre chez l'enseignant au moment où il évalue ses élèves. L'étude dégage deux grands facteurs: un facteur d'instruction, facteur d'impression générale traduisant un fort syncrétisme des jugements des comportements de participation, d'attitudes face au travail, de caractéristiques cognitives concernant les aspects convergents de la pensée: l'enfant jugé attentif, tenace et intelligent sera jugé positivement dans tous les domaines...; un second facteur de gestion du groupe-classe concerne plutôt les aspects sociaux du comportement: conformité à des règles sociales et morales de la vie scolaire.
"... les valeurs de la personne jugées les plus fondamentales dans l'appréhension d'autrui et qui gouvernent le syncrétisme de la perception de l'enseignant sont celles qui conditionnent, de son point de vue, la réalisation de ses objectifs professionnels et le degré de satisfaction de l'institution qui l'emploie".

On comprend bien à ce moment de l'analyse que le tri social fonctionne au niveau de l'acte quotidien sur le mode du conformisme des valeurs sociales dominantes, même si parfois le discours semble aller à leur opposé, ce qui rejoint la question de l'écart entre le dire et le faire en pédagogie. Ce conformisme véhicule comme valeur centrale l'idée d'une éducabilité différentielle: l'apprenant est plus ou moins apte aux études et au travers de la notation scolaire puis professionnelle cette aptitude

serait évaluée, aptitude dont le degré de possession justifie la position sociale. Bien sûr, il est aisé de montrer le caractère circulaire de l'explication dont les deux termes (aptitude et résultats) sont dans le même temps à la fois cause et effet l'un de l'autre: j'ai de bons résultats parce que je suis doué; je suis doué parce que j'ai de bon résultats !

On pourrait à partir de là basculer vers les thèses extrêmes, mais leur irréalisme freinerait d'emblée toute évolution. On peut réagir autrement, d'une manière plus positive car ouvrant les portes de l'action. Pour cela posons comme postulat de base, et tirons les conséquences de ce postulat en termes d'action, que l'être humain jouit d'une plasticité adaptative certaine, d'une capacité à s'auto-modifier en développant ses aptitudes lorsque le contexte physique l'y contraint et surtout lorsqu'il est stimulé dans ce sens par le contexte social.

Par exemple, le dépassement de l'orientation par l'échec réside dans l'introduction d'une nouvelle dimension éducative: l'éducation des choix.

C'est aussi une dimension éducative qu'essaie d'introduire l'évaluation formative qui est présentée comme une alternative à l'évaluation classique dont nous venons à grands pas de dessiner les plans de critique.

CHAPITRE 8

EVALUATION SOMMATIVE OU EVALUATION FORMATIVE ?

L'évaluation pédagogique, dans ses formes et ses procédures traditionnelles, ne résiste donc pas à un examen critique serré. Ce n'est pas pour autant qu'il faille "jeter le bébé avec l'eau du bain". En effet, le besoin d'évaluation est coalescent au besoin de reconnaissance sociale ressenti tout aussi bien par les écoliers que par les adultes en formation. De la même manière que les docimologues ont proposé des procédures statistiques permettant de progresser sur le plan formel de la quantification et de la mesure pour améliorer les défauts du système classique, les pédagogues et les psychopédagogues ont proposé un modèle d'évaluation qui soit plus pertinent techniquement et plus utile socialement. La principale innovation fut introduite par SCRIVEN qui distingue l'évaluation sommative de l'évaluation formative.

L'évaluation sommative intervient à la fin d'une unité d'enseignement ou d'une action éducative. Elle ressemble fort à l'évaluation améliorée par les techniques docimologiques sur le plan de la mesure et par la pédagogie par objectifs dérivée de BLOOM et ses élèves. L'évaluation sommative récapitule l'ensemble, ou plutôt une partie représentative d'un ensemble de connaissances, de comportements ou de savoir-faire dont les acquisitions ont donné lieu à un certain nombre d'heures d'enseignement. Elle permet d'un côté de situer les résultats de chacun des évalués par rapport aux normes attachées à un groupe-échantillon de référence, et de l'autre côté, par cumul des écarts individuels aux objectifs visés, d'apprécier la mesure des effets de l'intervention pédagogique. Nous retrouverons l'évaluation sommative quand nous essaierons de la mettre en relation avec l'évaluation formative. Cette dernière est nettement plus originale par rapport à l'examen classique dans la mesure où elle se déploie tout du long du procès d'éducation, pour assurer une fonction principale de régulation du processus d'apprentissage. Ainsi, et d'après DE LANDSHEERE, l'évaluation formative est celle qui intervient, en principe, au terme de chaque

tâche d'apprentissage, en ayant pour objet d'informer l'élève et le maître du degré de maîtrise atteint et éventuellement de découvrir où et en quoi un élève éprouve des difficultés d'apprentissage en vue de lui proposer ou de lui faire découvrir des stratégies qui lui permettent de progresser. Sont ainsi visés prises de conscience et redressements, individualisation de l'enseignement par l'ajustement de la démarche aux caractéristiques des publics... L'évaluation est donc qualifiée de formative qui apporte à l'intérieur du fonctionnement du système éducatif et de formation des informations utilisables pour adapter les activités d'apprentissage en fonction de chaque cas. Ainsi sont mises en oeuvre les améliorations nécessaires dès que s'en fait ressentir le besoin. Dans ce contexte il est remarquable que l'erreur ne constitue plus une faute répréhensible mais un indicateur du fonctionnement cognitif de chaque apprenant qui renseigne sur les étapes de l'apprentissage, le niveau et la qualité d'intégration des apports formatifs. C'est l'analyse fine de l'erreur qui aidera à déterminer les caractéristiques de la "remédiation" à entreprendre. Ainsi, la source de l'erreur pourra être référée soit à la démarche d'appropriation propre à l'élève -méthode inadaptée, manque de connaissances de base, choix inapproprié de la règle à appliquer,...-, soit à des données externes à l'élève -nature de la tâche, modalité langagière, contenus étrangers, contexte d'ambiance...-.

Pour CARDINET, la pratique de l'évaluation formative libère la démarche d'évaluation du formateur en l'affranchissant d'un nombre important de contraintes propres à l'examen: la limitation à la branche enseignée -toutes les sortes d'erreurs peuvent être prises en compte et on peut faire du français à l'occasion de la rédaction d'une solution de problème mathématique-; la limitation au programme de l'année -des enseignements hors programme éventuels peuvent être évalués; l'obligation d'arriver à une conclusion -si les circonstances sont défavorables, on peut différer les conclusions-; l'obligation de tout évaluer -l'évaluation formative peut porter sur le seul objectif de la leçon-; l'existence d'une norme de référence -des appréciations qualitatives choisies en rapport avec le travail fourni suffisent-; le choix d'un système de pondération -la différenciation de diverses catégories dans l'analyse des erreurs ne nécessite nullement une hiérarchisation de ces catégories-; l'estimation du niveau global atteint -ce n'est pas l'objet de ce type d'évaluation, ce qui ne doit pas empêcher une certaine appréhension des comportements dans une situation donnée-; le choix d'un seuil de suffisance -plus que le niveau en lui même, c'est la bonne direction qu'il s'agit de pointer-; la notation individuelle -on peut envisager quand les circonstances le commandent des retours d'informations collectifs-; les critères d'objectivité -la nature et le niveau des exigences sont adaptés à chacun-; le caractère favorable des conditions d'examen -il n'intervient pas en tant que condition de l'évaluation dans la mesure où de toutes les façons il y aura discussion autour des démarches et des résultats-; l'accord des examinateurs -a contrario, c'est le contact avec la diversité des points magistraux qui dans le cadre d'une pratique interactive va présenter un enrichissement supplémentaire-; l'équivalence des questions -là encore c'est la diversité des réponses à des situations plus ou moins éloignées car placées dans des contextes différents qui va enrichir l'observation-; la fiabilité des observations -la labilité d'un résultat n'est pas forcément une erreur d'évaluation mais une indication comportementale précieuse-. L'important, d'après cet auteur, reste bien que les données d'observation recueillies à l'occasion de l'évaluation puissent guider efficacement les interventions pédagogiques de remédiation à la condition évidente de

considérer que tout jugement évaluatif est daté et relativisé par un contexte. On retrouve une fois de plus pour le pédagogue et pour le psychologue de l'éducation la croyance dans le postulat de l'éducabilité de l'être humain.

On peut essayer de dégager les principes généraux d'une démarche d'évaluation formative à l'occasion de la correction d'un exercice qui relance l'apprentissage en apportant des informations supplémentaires à l'élève par des voies aussi variées que possible. Ces informations apportées vont concerner la matière, éventuellement la manière, et non pas l'élève en tant que personne, En effet, elles s'inscrivent dans un ensemble de boucles de régulation dont la finalité est l'amélioration des résultats produits par le système éducatif. Cet ensemble régulatif, d'après BONNIOL, assure trois fonctions principales :

1 – la reconnaissance des réussites;
2 – le repérage et la reconnaissance des erreurs;
3 – le réajustement des stratégies et la rectification.

Ces fonctions prendront valeur opérative à la condition d'avoir substitué une pédagogie par objectifs à la traditionnelle pédagogie du programme. Les objectifs choisis doivent être communiqués aux élèves afin qu'ils puissent se les approprier, les intérioriser et les intégrer à leur économie psychique et à leurs schèmes opératoires en tant que stimulateurs de l'investissement affectif et énergétique d'une part et en tant que guides de l'action d'autre part. Suffisamment difficiles pour représenter un défi à relever, les objectifs seront adaptés pour pouvoir être atteints sans trop de difficultés par l'ensemble des élèves. Dans ces conditions, la pédagogie par objectifs doublée d'une évaluation formative devient une source d'apprentissages efficace.

Bien entendu, une pratique convenable de l'évaluation formative passe par la prise d'un nombre minimum de précautions. Ainsi, il convient d'être prudent dans l'analyse de performances isolées et il est de loin préférable d'examiner une conduite globale afin de pouvoir comprendre la logique qui sous-tend la démarche de l'élève. En effet, la centration sur un point de détail pourra certainement se prolonger par la correction de quelqu'erreur ici et maintenant, mais sera insuffisante pour faire ressortir les invariants comportementaux, cognitifs et métacognitifs qui agissent comme générateurs permanents de l'erreur. Se centrer sur le résultat d'une addition de nombres décimaux amènera au mieux quelqu'apprentissage partiel, mais ne renseignera jamais sur l'origine affective d'une dyscalculie, ou sur la nature de la défaillance de l'appareil cognitif dans les processus déductifs... Il est pourtant essentiel de révéler la ou les sources des "pannes d'apprentissage" si l'on veut faciliter l'élaboration et la mise en place d'une réponse pédagogique ou psychopédagogique qui soit pertinente en termes de remédiation à long terme de toute une classe de problèmes ou de toute une catégorie de tâches. On comprend dans ces conditions la nécessité pour l'enseignant ou pour le formateur de tenir compte de l'ensemble des dimensions de la personnalité de l'élève ou du stagiaire afin qu'ils puissent se donner les moyens de comprendre les déterminants des attitudes, des conduites et des résultats habituels qu'ils ont pu observer. Cette approche "holistique" du sujet en situation d'apprentissage si d'évidence elle conditionne l'individualisation de l'action pédagogique pose le problème majeur de l'absence de formation psychologique approfondie des enseignants et des formateurs. Cette lacune dans la culture professionnelle de l'enseignant trace les limites de l'approche

individuelle et donc différentielle des difficultés d'apprentissage rencontrées par les élèves. Toutefois, cette limite se retrouvera déplacée quand le pédagogue pourra établir une relation de travail co-opératif avec un psychologue qui soit sensibilisé à l'intérêt de la psychopédagogie.

L'évaluation formative n'a rien à voir avec une pratique statique, routinière, ritualisée, itérative... C'est un processus dynamique qui demande à l'enseignant de déployer ses capacités d'analyse d'une manière adaptée à chaque cas. Ce processus s'inscrit dans un réseau d'interactions qui fait voyager l'information entre différents pôles: relation élève-situation pédagogique, relation élève-enseignant, relations inter-élèves,... La circulation de l'information dans ce réseau, lorsqu'elle est volontairement impulsée et contrôlée par le formateur, a pour objectif de provoquer les prises de conscience des fonctionnements cognitifs et affectifs chez les élèves et chez... l'enseignant qui sont utiles à la détermination des progrès dans les apprentissages.

En plus de ses capacités d'analyse, le formateur doit montrer et utiliser toutes ses qualités interprétatives. L'évaluation fait ressortir des faits par l'observation des conduites spontanées ou provoquées et par la mesure des degrés de réussite. Le fait tient du constat, du simple événementiel. Il ne possède aucune valeur explicative intrinsèque. Dire qu'à Marignan en 1515 s'est livrée une bataille, qu'en 1989 c'est Marseille qui a remporté la Coupe de France de football, que Jacques a eu 18/20 à sa dictée ne permet pas d'expliquer les causes et les conséquences de la bataille, de la victoire sportive, de la note d'orthographe. Pour que le fait évaluatif prenne sens il est nécessaire de le référer aux différents contextes pertinents par rapport au but poursuivi. Le problème n'est pas d'avoir à disposition un nombre élevé d'informations qui représenterait plutôt une gêne qu'une aide. Il est de savoir mettre en relation les données disponibles entre elles d'abord en respectant les critères de la cohérence interne, avec des éléments extrinsèques à la situation d'évaluation ensuite en respectant les critères de la cohérence externe. Le maniement des relations logiques est un moment utile au travail d'interprétation une fois que l'on a pu se donner une direction générale, un canevas, une pré-structure qu'il s'agit de fournir en contenus. Toutefois, au départ de la démarche interprétative, c'est l'intuition, la perception subjective, ou bien un travail de type "créativité" -brainstorming, libres-associations, mobilisation des acquis de l'expérience...- qui vont faciliter l'émergence du cadre et des principales hypothèses de l'interprétation. Le travail d'interprétation procède essentiellement d'un processus de rapprochement, de mise en relation, de comparaison entre ce l'on voit qui est de l'ordre du constat, de l'observation et ce que l'on sait qui est de l'ordre de la connaissance et de l'expérience professionnelles. Cette confrontation du percept et du concept qui accompagne l'appréhension du réel va ouvrir le plan de l'élaboration personnelle d'une explication, d'une compréhension, d'une reconstruction cognitive des caractéristiques des apprentissages de l'élève. Ainsi, la performance individuelle à une épreuve est rapprochée de la performance moyenne du même groupe d'âge; elle peut ensuite être mise en relation avec les performances obtenues aux contrôles précédents pour prendre en compte le progrès indépendamment du niveau.

Sur un plan plus analytique, et afin de prendre en compte dans l'analyse les relations qui se sont établies entre les caractéristiques de la tâche et la

performance de l'élève ou du formé, le travail d'interprétation peut passer en revue les sept paramètres dégagés par FEUERSTEIN pour caractériser un acte mental en les appliquant à l'épreuve évaluative (ces sept paramètres constituent la "carte cognitive") :

1 – Le contenu :

Le contenu de l'épreuve peut être dérivé de la géographie, de l'histoire, des mathématiques, de la physique, de la philosophie, de la technologie ou d'une combinaison de ces vastes disciplines où les outils d'une discipline peuvent être appliqués à une autre. Ainsi, les mathématiques peuvent être utilisées pour comprendre des données géologiques, pour réaliser l'implantation d'un bâtiment et des principes psychologiques peuvent expliquer des événements historiques, des faits économiques... Le contenu est l'un des domaines qui différencient largement les individus en fonction des acquis de leur expérience, de leur passé éducatif et culturel. Dans ces conditions, il faut se demander dans quelle mesure une performance traduit franchement un apprentissage scolaire lié à un progrès cognitif ou un certain degré de familiarité avec un contenu corrélatif d'une appartenance à un groupe social ou d'une histoire singulière. Un contenu inhabituel risque fort d'inhiber l'investissement dans le travail demandé, toute l'attention étant captée et mobilisée par la nouveauté. A l'opposé, un contenu trop familier aura un effet identique de désinvestisement lié cette fois-ci au désintérêt porté aux choses trop familières.

2 – Les opérations mentales :

Si la question des contenus se rapproche plutôt des aspects didactiques, la question des opérations mentales en jeu dans les apprentissages est plutôt reliée à la psychologie cognitive et en règle générale pose des problèmes de... familiarité aux enseignants et aux formateurs. L'opération mentale est définie par PIAGET et INHELDER comme un ensemble d'actions intériorisées, organisées et coordonnées avec lequel nous pouvons élaborer, transformer d'une manière cognitive les informations parvenant des sources externes et des sources internes. Les opérations peuvent s'ordonner de la simple reconnaissance et identification des objets à des activités plus complexes telles que la classification, la sériation, la multiplication logique,... Les opérations peuvent s'appliquer à des informations existantes ou nécessitent l'élaboration préalable de nouvelles données comme dans le cas du syllogisme, de l'analogie, du raisonnement par inférence...

Le repérage par le formateur des opérations mentales nécessaires à la réalisation d'un exercice est de toute première importance. En effet, une difficulté d'apprentissage observée de l'extérieur peut très bien être liée à l'absence de maîtrise d'une opération mentale. Tant que la difficulté cognitive ne sera pas repérée et traitée, l'assimilation du contenu scolaire butera sur ce manque d'outils intellectuels.

(Remarque: la psychopédagogie cognitive propose pour les pré-adolescents, les adolescents et les adultes des programmes de "remédiation cognitive" de plus en plus performants).

3 - La modalité langagière :

La pensée peut trouver son expression par sa traduction dans différents types de langages. Une même idée peut se "matérialiser" dans des mots - modalité verbale -, dans des nombres - modalité numérique -, dans des symboles - modalité symbolique -, dans des dessins - modalité figurative -, dans des schémas, des graphiques, des tableaux... - modalité graphique -... ou dans la combinaison de deux ou plusieurs modalités. Il existe de multiples formes de codage, de la mimique et autres modes de communication métalinguistique aux signes conventionnels qui fonctionnent de manière indépendante des contenus qu'ils véhiculent.

Globalement, c'est le contexte culturel qui détermine la ou les modalités de communication les plus usuelles, mais à l'intérieur d'une même culture, on remarque des affinités individuelles marquées dans le choix spontané des "signifiants", ou dans le degré de compréhension et de performance selon le signifiant imposé. Ainsi, il n'est pas rare d'observer qu'un élève qui réussit fort bien des exercices mathématiques présentés dans une modalité symbolico-numérique va échouer lorsque le même type d'exercice est présenté uniquement par un texte. Il en résulte que les modalités choisies pour l'exercice d'évaluation ne sont pas neutres et influent directement sur le niveau de réussite de l'élève ou du stagiaire. En conséquence, on ne peut pas conclure à une incapacité intellectuelle définitive tant que l'ensemble des possibilités de présentation de l'exercice n'aura pas été épuisé. D'autre part, on voit tout l'intérêt qu'il y a à connaitre la modalité préférée de chacun des apprenants. Elle pourra être volontairement choisie comme contenant pour faciliter tel ou tel apprentissage, comme détour pour surmonter une difficulté apparue dans une autre modalité, comme point d'appui pour favoriser les transferts et les généralisations des acquis scolaires.

4 - La phase de l'acte mental :

Un acte mental peut être grossièrement divisé en trois phases: l'entrée ou prise d'informations, l'élaboration ou transformation des données, la sortie ou formulation de la réponse. Bien entendu ces phases sont reliées entre elles et le rôle joué par chacune ne peut être considéré qu'en relation avec celui des deux autres. Toutefois, pour mieux appréhender le rapport fonctionnel qui s'établit entre un élève et une situation-problème, il est méthodologiquement intéressant de distinguer entre elles. Tous les cas deviennent possibles pour expliquer la raison d'un échec ou d'une difficulté. Il se peut que l'élève ou le stagiaire n'ait pas, pour une cause qu'il resterait à déterminer, pris le soin de percevoir et d'enregistrer l'ensemble des éléments du problème à résoudre ou de la tâche à exécuter. Un exemple fréquent en mathématique est celui du problème correctement raisonné, mais faux car les calculs numériques ont porté sur des nombres autres que ceux donnés dans l'énoncé: il y a eu défaillance au moment de l'enregistrement perceptif d'un ou de plusieurs nombres. Il se peut également que l'apprenant ait correctement saisi les informations pertinentes, qu'il les ait correctement transformées, mais qu'il se trompe dans la matérialisation de cette réponse: il ne coche pas la bonne case du Q.C.M., il oublie de recopier une page entière de son brouillon de rédaction, il noie sa réponse dans un flot d'éléments inutiles ou superflus... Il se peut enfin que l'erreur soit la

conséquence d'une insuffisance dans l'élaboration cognitive proprement dite, soit que l'élève ne déroule pas un raisonnement correct -choix inapproprié-, soit qu'il présente des lacunes dans son répertoire de schèmes de résolution disponibles.

L'évidence de l'intérêt de ces distinctions pour l'exercice de l'évaluation formative saute aux yeux: en "localisant la panne" on peut également localiser l'intervention éducative et moduler son objectif et ses moyens en conséquence.

5 - Le niveau de complexité :

Le niveau de complexité peut être compris comme la quantité et la qualité des unités d'informations nécessaires pour produire un acte mental donné. Le nombre d'unités peut concerner des objets isolés ou des groupes d'objets, mais l'accent est placé sur les unités et/ou les opérations. Bien que difficile à définir, une unité comprend les dimensions qui sont d'une immédiate nécessité pour qu'un acte mental spécifique prenne place. La complexité demande une appréciation différenciée qui considère simultanément le nombre d'unités d'information et la qualité de ces unités en termes de degré de nouveauté ou de familiarité. Plus les unités sont familières, moins l'acte mental est complexe, même si les unités sont multiples; inversement, moins elles sont familières et plus l'acte est complexe.

Cette conception de la complexité transposée dans le domaine de l'évaluation formative a le mérite de relativiser un comptage objectif (le nombre d'unités d'information) par une perception subjective (l'impression de familiarité ou de nouveauté). Cette relativisation met à l'abri d'une interprétation brutale ou entière en réintroduisant le point de vue de l'évalué qui est lui seul en mesure de formuler l'impression de familiarité ou de nouveauté. Les prolongements de cette manière de voir les choses dans la régulation de l'action éducative sont de première importance: ils concernent directement les conditions de la construction d'un matériel pédagogique qui puisse faciliter la réussite en évitant de placer les élèves ou les stagiaires face à un flot d'informations et/ou face à des informations qui leur sont étrangères.

6 - Le niveau d'abstraction :

A l'aide du niveau d'abstraction se définit la distance entre l'acte mental donné et l'objet ou l'événement sur lequel il opère. Les paramètres décrits supra peuvent se concevoir comme opérant directement sur les objets ou les événements. Ainsi, le contenu peut comprendre des objets qui peuvent être perçus d'une manière sensorielle ou appréhendés complètement par une manipulation motrice. Différentes modalités peuvent être utilisées et des opérations comme additionner, soustraire, grouper des objets réels ou leurs substituts appliquées. Toutefois, les paramètres discutés précédemment ne prennent pas en compte l'étendue qui sépare les éléments sur lesquels les opérations sont utilisées et les objets ou événements perçus à l'origine.

Ainsi, un acte mental peut comprendre des opérations sur des objets eux-mêmes comme les assortir en fonction d'un système de classification. A un niveau

d'abstraction plus élevé, les opérations peuvent intervenir sur des représentations des objets, comme l'application d'une classification à un ensemble d'objets hypothétiques. Encore plus abstraite est l'application d'opérations à des propositions purement hypothétiques, sans référence à des objets ou événements réels ou imaginés. Ainsi, une hiérarchie de niveaux d'abstraction peut être établie, utilisant comme critère la distance entre l'opération mentale et l'univers des objets ou des événements auxquels elle est appliquée.

La société en général et l'école en particulier fondent l'une des dichotomies sociales de base entre les travailleurs et les intellectuels sur le degré de maîtrise de la capacité de raisonner dans l'abstrait. Trop souvent cette capacité est vue comme une aptitude naturelle qui va justifier la stratification sociale, le tri scolaire susceptible de l'alimenter par la différenciation des cursus scolaires en cycles d'enseignement long pour "les élèves abstraits" et cycles d'enseignement court pour "les élèves concrets". Une évaluation formative digne de ce nom, c'est-à-dire une évaluation pédagogique dont la fonction essentielle ne serait pas le classement mais la détermination des conditions et des moyens de faire progresser les élèves au maximum, ne peut pas se cantonner dans cette position statique et se doit de refuser le glissement du "on est abstrait" au "on naît abstrait". Le degré d'abstraction maîtrisé par chacun des élèves ou des stagiaires est bien évidemment une donnée essentielle à apprécier. La connaissance de cette donnée est utile de plusieurs points de vue. D'abord elle éclaire sur le sens de certains échecs ou de certaines réussites. Secondement elle permet d'ajuster le niveau d'abstraction des exercices aux capacités réelles des élèves et évite en conséquence de les placer face à des situations qu'ils ne sont pas en mesure de maîtriser. Troisièmement, et c'est là le principal, cette connaissance, pour les élèves dont les capacités d'abstraction sont réduites, facilite l'élaboration et la mise en oeuvre d'exercices spécifiques ayant pour but de développer leur capacité défaillante, de surmonter cette défaillance jusqu'à atteindre un niveau suffisant pour progresser scolairement et professionnellement selon l'attente des professeurs, des formateurs, de l'encadrement...

7 - Le niveau d'efficience :

L'acte mental peut aussi être décrit et analysé en fonction du degré et du niveau d'efficience avec lequel il est produit. Comme critère de l'efficience on peut utiliser le rapport rapidité-précision et/ou la quantité d'effort objectivement et subjectivement exercé par l'individu dans la production de l'acte considéré. Le concept d'efficience peut se concevoir comme une dimension qui est différente, à la fois qualitativement et quantitativement, des six autres paramètres bien qu'elle puisse être déterminée ou influencée par l'un ou l'autre des paramètres ou par leur combinaison. Par exemple, un bas niveau de complexité, résultant de la familiarité avec un contenu donné ou avec des opérations cognitives particulières, peut rendre efficient l'acte mental qui opère sur ces domaines particuliers.

Le manque d'efficience, défini par la lenteur, une production réduite, ou par l'imprécision, le traitement inefficace d'un problème, peut être sans aucun rapport avec la capacité d'un individu de se saisir et de résoudre un problème particulier. Notre incapacité d'isoler l'efficience de la capacité est une importante source

d'erreur dans l'évaluation, non seulement de la capacité individuelle réelle mais même de son répertoire d'information et d'outils. Une telle erreur, provenant de la confusion entre l'efficience et la capacité, résulte dans l'appellation incorrecte et la formulation d'un pronostic basé sur le postulat erroné d'une faible capacité qui a été dérivée seulement de l'observation d'un bas niveau d'efficience. Cela, au passage, détermine le bas niveau des objectifs d'éducation, d'orientation, et autres établis pour l'individu.

L'appréhension efficace d'une tâche est toujours relative à une variété de facteurs intrinsèques et extrinsèques à la tâche qui affectent l'individu. L'anxiété ou le manque de motivation peuvent interrompre l'efficacité individuelle dans une tâche particulière. Toutefois, le degré de cristallisation de l'acte mental et son automatisation peuvent déterminer la nature et l'aptitude de l'impact de ses variables sur l'efficience de l'individu. Plus les modèles de comportement sont récemment acquis, plus vulnérables et moins résistants ils seront face à l'impact de facteurs interférant. Plus le modèle est installé et cristallisé et moins les facteurs émotionnels vont interrompre son propre fonctionnement.

Il en est de même pour la motivation. En décrivant la nature et la quantité de la motivation de groupes d'enfants présentant différents niveaux de réussite, on est tenté d'expliquer les bas niveaux de fonctionnement par le concept de "motivation", qui est considéré comme étant différent et de manière concomitante plus bas chez les enfants à basse performance. Toutefois, comme l'efficience individuelle est déterminée, au moins en partie, par sa familiarité avec les objets, les événements et les opérations sous-jacents, le degré de motivation nécessaire pour mobiliser l'énergie demandée pour produire un acte mental est plus élevé chez les élèves démunis culturellement que chez les plus favorisés. L'investissement différentiel, déterminé par les différents niveaux d'efficience caractéristiques de ces groupes, déterminera lui-même la somme de motivation requise pour produire l'investissement nécessaire. Cependant, même si la motivation est maintenue égale, le résultat d'un acte mental peut différer puisque les niveaux d'efficience les plus bas demandent des sommes d'efforts plus grandes.

L'une des plus importantes dimensions du développement de l'efficience des opérations cognitives est l'automatisation progressive de certains comportements, un processus qui amène une réduction de la quantité et du niveau de l'effort nécessaire pour produire un acte mental donné. Le processus d'automatisation, à travers la familiarisation avec certains composants de la tâche, provoque une possibilité de consacrer l'essentiel de l'énergie dans les zones qui demandent encore de l'investissement pour la maîtrise de la tâche d'une manière rapide et précise. Moins l'opération est automatique, plus grande est la vigilance requise pour maîtriser l'ensemble de ses composants et moins efficient sera le traitement de la tâche. De même, l'inefficience peut être produite dans une tâche familière si un essai est tenté pour découvrir dans la tâche des relations inemployées, parce que l'effort porte sur des zones autres que celles pour lesquelles la tâche était initialement prévue. On peut observer une baisse dans le fonctionnement d'individus hautement intelligents quand ils sont confrontés à des tâches trop faciles pour eux parce qu'ils investissent beaucoup plus qu'ils n'en ont besoin en recherchant des éléments plus complexes que ceux proposés.

Quand des opérations et des stratégies nouvelles acquises sont évaluées, il faut considérer le haut niveau d'investissement nécessaire à la première étape de leur utilisation. Corrélativement, une pondération doit être accordée pour une basse efficience individuelle, qui est attribuable à l'aspect récent de l'acquisition et au manque d'automatisation des outils cognitifs utilisés pour la solution d'un problème donné.

"En analysant un acte mental, la raison de l'inefficience peut être localisée et associée avec l'un des six paramètres préalablement mentionnés: le contenu; le niveau de complexité, qui est défini par le nombre d'unités d'informations corrigé par le degré de nouveauté; le niveau d'abstraction, qui demande l'organisation des données, l'établissement de relations, l'organisation et la catégorisation des relations elles-mêmes; et le degré de familiarité de l'opération, ou la quantité différentielle d'investissement nécessaire pour décoder une modalité spécifique de présentation du problème".

Nous avons longuement donné la parole à FEUERSTEIN pour expliquer le dernier point de son système d'analyse de l'acte mental, point qui concerne l'interprétation si délicate et trop souvent inexistante du rapport entre la performance et la capacité. L'assimilation rapide de la seconde à la première empêche de poser clairement les questions de la motivation, de l'investissement dans la tâche liées à la représentation de lui-même dans la situation que se construit l'élève, et aussi de la tonalité affective générale... S'il est vrai que la réussite traduit l'existence de la capacité, la réciproque n'est pas justifiée, comme nous l'a montré FEUERSTEIN. Pour le formateur qui pratique l'évaluation formative, il est donc tout à fait essentiel d'analyser les erreurs dans toutes leurs dimensions, cognitives, métacognitives, psycho-affectives... C'est bien à partir du moment où il aura pu sérier l'ensemble des causes d'erreurs possibles et déterminer celles qui ont la plus forte probabilité que le formateur peut aller plus loin dans la mise en place de son dispositif évaluatif pour vérifier ses hypothèses et finalement décider des exercices ou des dialogues à proposer pour apporter des améliorations sensibles. Toutefois, il est certain qu'une approche psychologique globale des difficultés d'apprentissage ne fait pas encore franchement partie du répertoire des comportements du maître. Dans cette optique, il est fréquent que les problèmes de motivation, d'investissement, d'affectivité... ne trouvent pas de solution pédagogique satisfaisante et il est nécessaire de lancer des recherches qui puissent apporter les informations pertinentes aux enseignants et aux formateurs.

Démarche très exigente donc que celle d'une évaluation formative qui se voudrait efficace pour les élèves en visant la transformation de leurs représentations, en renforçant leurs motivations internes ou externes, en leur transmettant savoirs et savoir-faire à la fois utiles à leur épanouissement individuel, à la résussite de leur formation et à leur trajectoire sociale.

On voit bien que ce qui marque à la fois l'originalité et la difficulté de l'évaluation formative, c'est la perspective diagnostique, détour analytique nécessaire pour découvrir l'intervention en retour la plus efficace, toutes les productions, comportements, verbalisations, réactions et démarches de l'élève, ou de son entourage, pouvant suggérer des hypothèses à vérifier. Au fond, une évaluation

formative réussie est celle qui produit des jugements de valeur sur des réussites objectivées et qui recherche les causes des échecs pour définir l'intervention qui va amener la réussite et le jugement positif.

Cette pratique de l'évaluation peut se dérouler en continu, dans le but de nourrir la régulation positive des apprentissages. Elle nécessite de la part du maître ou du formateur l'inébranlable volonté de toujours rechercher causes et solutions à l'échec, et de ne jamais laisser un élève ou un stagiaire enfermé dans une difficulté. Le pédagogue doit rester convaincu qu'il existe toujours un moyen pour stimuler une dynamique de progrès et de dépassement dans le processus de formation de l'enfant ou de l'adulte. Certes le chemin reste étroit qui consiste à exercer d'éventuelles pressions éducatives sans pour autant projeter le schéma d'une progression didactiquement linéaire en place d'un développement psychologique différentiel qui suit ses propres lois. Il ne faut pas oublier que le "moteur" du développement et de l'apprentissage est un moteur interne et que de l'extérieur un pédagogue ne peut espérer exercer un pouvoir de transformation que dans l'aménagement de l'environnement, dans "l'écologie" du système éducatif, et principalement dans la modulation de la relation éducative. Comme le souligne CARDINET,
"La seule référence possible pour une évaluation éducative est alors une direction de développement. Sera positive et à soutenir toute transformation de l'individu qui élargit sa prise de conscience du monde et accroît ses possibilités d'action... On pourrait appeler évaluation formative continue cette prise d'information qui accompagne un être en développement et s'efforce de faciliter à chaque moment sa croissance, en aménageant au mieux son environnement".

L'enseignant ou le formateur étant un sous-ensemble de la situation éducative globale, il devra exercer ce contrôle continu sur ses propres actes, gestes et paroles. L'auto-évaluation par le pédagogue -ou dans les cas plus favorables par l'équipe psycho-pédagogique- de la démarche qu'il met en oeuvre est un corrolaire d'une évaluation formative intégrale. Dit autrement, la formation en continu des élèves ou des stagiaires passe par la formation en continu des maîtres ou des formateurs. On encore, puisqu'il se préoccupe de transformer les autres, le pédagogue doit également se préoccuper des conditions et des processus de sa propre transformation, essentiellement en prenant conscience un maximum de ce que lui apprennent ses élèves.

Sur le plan pratique, on peut commencer à s'exercer à l'évaluation formative de manière toute simple, en demandant à l'apprenant qui éprouve une difficulté de le signaler très vite. Le premier niveau d'intervention consiste pour le formateur dans l'aide à apporter à l'élève pour qu'il puisse identifier sa difficulté et la verbaliser le plus clairement possible. Cette attitude permet de constater que cette intervention facilite les prises de conscience par l'élève du problème posé, de la situation à résoudre, des données pertinentes à considérer, des informations manquantes, des opérations nécessaires, du choix d'une stratégie de résolution, du niveau de difficulté et de l'investissement à mobiliser... L'enseignant s'aperçoit alors de l'immense richesse et de la diversité des modes de fonctionnement individuels, et c'est en respectant cette variété que le pédagogue peut le mieux ajuster sa réponse aux besoins de chacun. L'habitude de "penser" ses difficultés

amène rapidement chez l'élève une attitude d'auto-contrôle, que l'on retrouve dans les processus de fabrication industrielle sous le terme de "qualité totale". Le formateur constatera alors que le seul fait de s'interroger sur le "comment" d'une résolution de problème amène le formé a trouvé par lui-même nombre de solutions, tant il est vrai que la facilité de résolution est d'abord fonction de la manière de considérer la tâche. Il est bien sûr souhaitable de multiplier les occasions et de diversifier les partenaires avec lesquels le formé va pouvoir échanger et obtenir des retours d'information sur ce qu'il vient de faire, sur la démarche qu'il a préféré utiliser... La mise en commun, la discussion collective où la voix du maître est une parmi d'autres, la "socialisation" des démarches individuelles sont autant de moments privilégiés pour acquérir une capacité de décentration critique vis à vis de ses productions et comportements et vis à vis de ceux des autres. De la sorte, tout en facilitant l'acquisition de connaissances spécifiques à une matière ou à un domaine professionnelle, l'évaluation formative participe à la formation globale de la personne en transmettant progressivement les moyens de l'auto-formation et de l'auto-évaluation.

Il est possible de "techniciser" l'évaluation formative, en prenant garde, bien évidemment, à ne pas systématiser trop fortement ce qui deviendrait d'autres rituels, à ne pas s'enfermer dans des cadres et des procédures rigidement itératives, à ne pas atomiser l'information sans mettre en relation les unités, à ne pas élaborer des systèmes fermés qui ne font que s'expliquer eux-mêmes et qui stérilisent toute créativité et spontanéité... Rendre plus technique l'évaluation formative peut se faire en procèdant à trois séries de différenciations: les moments, les niveaux, les objectifs. Pour chaque plan de différenciation, c'est en premier lieu l'objet de l'évaluation qui va changer.

On peut distinguer ainsi les principaux moments du processus d'évaluation. L'évaluation du début d'une unité d'enseignement ou de formation va permettre de faire le point sur l'état de préparation des élèves, sur l'ensemble de leurs dispositions, sur les requis cognitifs nécessaires aux prochaines acquisitions... Il s'agit en quelque sorte de s'assurer de la probabilité de réussite de l'élève (on dit alors de l'évaluation qu'elle a une visée pronostique) et si nécessaire d'apporter les correctifs adéquats (travail pédagogique en termes de prévention). En cours d'apprentissage, l'évaluation aura plutôt pour objet la régulation des acquisitions. En cas de difficulté, elle devient franchement diagnostique et a pour objet la définition de la nature des difficultés, des blocages et des manques, ainsi que la recherche des moyens d'intervention les mieux adaptés au problème posé (travail psychopédagogique de remédiation). L'évaluation formative intervient également au moment où l'on fait le point d'étape intermédiaire dans un procès d'apprentissage: on peut faire la part de ce qui est effectivement acquis et de ce qui reste à acquérir, reprendre et moduler les stratégies et méthodes de travail jusque-là utilisées pour les rendre plus efficaces,... En fin d'apprentissage, l'évaluation à caractère formatif peut une dernière trois intervenir pour s'intégrer dans le bilan final et aider à tirer l'ensemble des enseignements de l'expérience d'apprentissage.

Le second plan de différenciation touche à la question du niveau auquel se fixe l'évaluation. S'intéresse-t-on à la réponse, ou aux démarches de résolution cognitives mises en oeuvre, ou aux capacités mises en oeuvre, ou aux capacités

métacognitives qui accompagnent les démarches, ou aux représentations intimes de l'élève qui s'élaborent ou se mobilisent face à l'exercice, ou aux aspects de l'économie affective de l'élève...? Chaque plan a son importance et a à être traité à un moment ou à un autre, le plus souvent en relation avec un autre. Ainsi, on oppose souvent la centration sur le résultat et la centration sur la démarche. Les deux sont nécessaires. Considérer seulement le résultat selon l'adage "il n'y a que le résultat qui compte" ne fournit pas à l'enseignant d'indications pour améliorer le résultat de ceux qui en ont besoin. De plus cela exaspère les aspects compétitifs des systèmes de formation et renforce "l'inégalité des chances". A contrario, considérer uniquement des démarches sans se soucier de leurs résultats effectifs peut engendrer de sérieuses déresponsabilisations, les élèves perdant l'habitude de faire le lien entre la cause (leur démarche) et l'effet (une réussite ou un échec).

Enfin, rendre plus technique l'évaluation formative passe par la mise en oeuvre d'une pédagogie par objectifs. Beaucoup a été écrit à ce sujet, à la suite des propositions de BLOOM et notre intention est bien de rester sur le plan de l'évaluation. Disons simplement que, théoriquement, la découpe du domaine à enseigner en plusieurs objectifs correctement définis en termes de connaissances et/ou de comportements objectivables permet d'élaborer le choix des moyens didactiques à mettre en oeuvre pour atteindre ces objectifs ainsi que les procédures évaluatives appropriées pour réguler les apprentissages: on parle alors parfois d'évaluation de maîtrise.

Que l'on aille plus ou moins loin dans la technicité, il faut bien considérer la conséquence essentielle de l'évaluation formative sur les pratiques classiques: nous voulons évoquer une certaine forme d'individualisation de l'enseignement et de la formation. En effet, la compréhension progressive du fait fondamental du particularisme des apprentissages amène à prendre en compte chacun d'eux. Il est temps alors d'adapter le rythme de la progression, d'ajuster les exigences, et surtout de comprendre la logique des conduites de chacun pour étayer l'intervention d'aide pédagogique sur l'existant. Et individualiser efficacement, tout en utilisant les phases de travail coopératif ou d'échanges sur les stratégies afin de participer au dépassement des égocentrismes en visant un enrichissement mutuel, va à l'encontre de trois siècles d'enseignement collectif. Pourtant, dans l'idéal, on pourrait imaginer que l'"individualisation socialisée" se prolonge par une connaissance beaucoup mieux approfondie des apprenants qui puisse alimenter d'informations en retour les autorités scolaires et favoriser ainsi gestion de l'école et adaptation du système scolaire.

La tentation pourrait exister d'opposer d'une manière caricaturale l'évaluation formative à l'évaluation sommative, la première pouvant être rejetée comme trop moderne, trop éloignée des contraintes réelles ou perçues comme telles par des enseignants plutôt enclins à défendre les aspects évaluatifs traditionnels, la seconde pouvant éveiller les rétiscences de pédagogues progressistes qui se contenteraient de l'assimiler uniquement aux procédures de contrôle classiques. L'analyse fait apparaître un paysage plus subtil qu'une simple opposition. L'évaluation formative et l'évaluation sommative sont comme les deux faces d'une même pièce de monnaie: différentes dans leurs apparences, elles n'en appartiennent pas moins au même objet. Toute évaluation sommative se réfère, même implicitement, à la

face formative dans la mesure où elle permet à l'évalué de savoir avec précision ce qu'il sait et ce qu'il ne sait pas. Réciproquement, toute évaluation formative se réfère ne serait-ce qu'implicitement à la face sommative du seul fait de la répétition et de la mise en relation des évaluations successives. L'évaluation sommative se présente de deux manières. La première, que l'on connaît bien, correspond à la forme des examens traditionnels de fin d'année, globaux ou sous forme d'unités capitalisables: il s'agit de l'évaluation sommative externe qui débouche sur la reconnaissance sociale en termes de passage dans la classe supérieure ou de délivrance d'un diplôme... La critique de cette évaluation a fait l'objet des chapitres précédents; nul besoin d'y revenir. On peut aussi envisager une évaluation sommative interne, voire auto-pratiquée, qui consiste à totaliser les apprentissages effectivement réalisés à la fin d'une unité de travail, parfois d'une unique séance d'apprentissage. L'ensemble de ces évaluations partielles peut réaliser une forme de contrôle continue des connaissances, chacune d'entre elles renseignant sur le degré d'intégration d'une certaine unité didactique, ou le degré d'atteinte d'un objectif particulier.

Une autre distinction est apparue qui différencie l'évaluation sommative critériée de l'évaluation sommative normée. La seconde, dite encore évaluation de jugement relatif, se rapproche des tests de performances (voir chapitre suivant) qui consistent à attribuer un rang à chaque élève ou stagiaire en fonction du résultat quantitatif qu'il obtient, comparativement à l'ensemble des performances obtenues par ses pairs confrontés à la même tâche. La première, dite également évaluation de maîtrise ou encore évaluation par objectif ou évaluation de jugement absolu, est beaucoup plus exigeante sur le plan qualitatif puisqu'elle doit fournir des indications précises sur la réalisation d'un critère de réussite à une performance: Untel, qui fréquente une formation de mécanicien-auto sait changer tous les embrayages de toutes les voitures; Untel qui est resté une année en Grande-Bretagne comprend tous les accents anglais; Untel, qui est au cours élémentaire première année sait calculer les soustractions des nombres entiers avec retenues... On voit que dans un cas la référence formelle est de l'ordre du comptage et de la statistique alors que dans l'autre elle est fondée sur l'observation directe de comportements, de l'utilisation des savoirs et des savoir-faire. L'évaluation normée répond plutôt aux besoins de classement, dans les situations de sélection par exemple. L'évaluation critériée, elle, répondant plutôt aux besoins de bilan analytique des connaissances et/ou d'inventaire de comportements, devrait trouver place dans l'élaboration de profils pédagogiques d'un individu ou d'une classe, dans les procédures de régulation des apprentissages, dans l'élaboration des programmes d'enseignement, dans l'élaboration des listes de compétences professionnelles...

L'évaluation formative et l'évaluation sommative critériée proposent aux enseignants, aux formateurs et aux administrateurs des pistes et des moyens concrets d'amélioration des procédures d'évaluation classique. Toutefois, ces alternatives ne se développent que fort lentement et ont subi à leur tour toute une série de critiques.

Par exemple, pour BARBIER qui pose le problème sur un plan social élargi, la distinction entre évaluation sommative et évaluation formative n'est pas assez approfondie et elle relèverait d'une attitude "scolaro-centrique", dans la mesure où

elle ne s'est pas suffisamment détachée des mécanismes sociaux qui tendent à autonomiser l'évaluation dans ses aspects techniques. Trop axée sur le moment de l'évaluation par rapport au déroulement de l'acte de formation comme forme instituée, elle n'aurait pu produire une véritable articulation entre la réflexion sur la nature de la formation, la réflexion sur les outils de l'évaluation, et celle sur ses fonctions.

Sur un plan technique, une évaluation formative strictement attachée au concept d'objectif deviendrait vite ingérable: par exemple l'enseignement du français à l'école primaire nécessiterait la mise en oeuvre de sept cents instruments d'évaluation d'après le programme canadien SAGE présenté par DUSSAULT. Si chacune des disciplines enseignées nécessitent autant d'instruments, alors le contrôle prendra définitivement le pas sur l'enseignement. En plus de l'aspect gestion du temps, l'atomisation en objectifs trop étroits se révèle contraire à la dynamique des apprentissages dans leurs aspects fonctionnels. On n'apprend pas à lire lettre après lettre, mais en mettant des lettres en relation entre elles, avec des sons, avec des images, avec des sensations et des significations... Dans ces conditions, contrôler une connaissance ponctuelle ne renseigne plus du tout ou que d'une manière limitée et aléatoire sur l'aspect général de l'apprentissage: connaître l'alphabet ne garantit pas la capacité de compréhension d'un texte écrit... En quelque sorte, le découpage en objectifs ignorerait l'aspect dynamique de tout apprentissage: l'évaluation indique si la réponse produite est conforme au modèle de réponse attendue, mais ne renseigne pas sur l'essentiel, à savoir le processus d'élaboration de cette réponse, seule garantie de la maîtrise réelle et donc de la reproductibilité de la conduite. Le découpage d'un objectif global en de trop nombreux sous-objectifs dilue le sens général de l'apprentissage du fait de la difficulté pour les élèves de rétablir les relations entre les parties dont ils ne connaissent pas la logique qui à présider à leur délimitation.

Il y aurait donc intérêt à distinguer entre une évaluation par objectifs au caractère formel qui ne fait que segmenter la matière tout en restant centrée classiquement sur la justesse du résultat et une véritable évaluation formative qui déplace le centre de l'observation vers les fonctions cognitives et les opérations mentales générales à mettre en oeuvre pour penser efficacement et réussir toutes sortes de familles de problèmes.

C'est peut-être la difficulté de ce recentrage qui explique que même suite à une formation spécifique à l'évaluation formative nombre d'enseignants restent quasi englués dans leurs habitudes évaluatives premières. J.J. BONNIOL remarque cet échec au moment des conseils de classe où dans les meilleurs des cas on fera le point sur chacun des élèves, mais en restant au niveau du constat et du jugement de valeur et sans jamais amorcer un quelconque "plan de soin" :
"La confusion persiste, soutenue par les logiques secondes qui prennent le rôle principal; il s'agit bien sûr des logiques de la didactique, de la discipline, de l'organisation pédagogique ou administrative, qui s'imposent au nom de la cohérence, du droit à la diversité, du rythme qu'il faut tenir, du niveau qu'il faut garantir dans la discipline; bref au nom du nécessaire qui n'est jamais suffisant et qui n'en est pas mieux garanti, puisque les boucles de régulation qui lui donnerait du sens ne fonctionnent pas".

Si tant est que l'évaluation formative représente une alternative potentielle valable pour surmonter les insuffisances et les effets négatifs de l'évaluation pédagogique traditionnelle, les quelques expériences de sa diffusion à l'intérieur de l'institution montrent, une fois de plus, les difficultés à enraciner une innovation dans le corps des pratiques sociales. Nous pensons que l'apprentissage de la passation des tests pédagométriques par les enseignants et les formateurs peut constituer une phase intermédiaire intéressante et efficace pour amorcer le décentrage nécessaire à la réussite de toute évolution.

CHAPITRE 9

QU'EST-CE QU'UN TEST ?

Avant de décrire l'utilisation des tests en pédagogie comme étape de l'amélioration des pratiques évaluatives, il convient d'essayer de comprendre ce qu'est un test et comment il fonctionne. En effet, l'usage des tests n'est pas apparu d'abord en pédagogie, mais en psychologie.

Etymologiquement, le mot test est dérivé du latin testum qui veut dire vase de terre. Dans l'ancien français, le mot têt désignait un pot de terre servant à l'essai de l'or en alchimie. Puis il est emprunté par l'anglais où il signifie "examen, épreuve".

C'est McK. CATTELL qui introduit le terme dans le domaine de la psychologie en 1890 dans un article de la revue *MIND* où il décrit sous le nom de "mental tests" des épreuves standardisées. Les psychologues BINET et HENRI font repasser le mot dans la langue française dès 1895.

L'Association internationale de Psychotechnique a adopté la définition suivante du mot test: "C'est une épreuve définie, impliquant une tâche à remplir, identique pour tous les sujets examinés, avec une technique précise pour l'appréciation du succès ou de l'échec, ou pour la notation numérique de la réussite. La tâche peut comporter une mise en oeuvre, soit de connaissances acquises **(test pédagogique)**, soit de fonctions sensori-motrices ou mentales **(test psychologique)**".

Plusieurs systèmes de classification des tests ont pu être proposés eu égard à leur grande diversité.

On distingue les tests individuels où l'examen concerne un seul sujet à la fois des tests collectifs où la passation s'applique à un groupe de sujets plus ou moins nombreux.

Les tests "papier-crayon", comme leur nom l'indique, ne nécessitent qu'une feuille d'interrogation et un crayon alors que d'autres tests demandent l'utilisation d'un matériel plus ou moins sophistiqué, des simples cubes que le sujet doit combiner à des dispositifs électro-mécaniques ou électroniques complexes dans lesquels le sujet actionne des commandes.

Les "tests verbaux" qui utilisent principalement la langue comme support de communication et d'élaboration se différencient des tests non-verbaux qui font appel à des modalités d'expression numérique, spatiale, symbolique, concrète...

Dans certains cas, le sujet resté libre de formuler sa réponse comme bon lui semble, c'est lui qui l'élabore. A contrario, dans les questionnaires à choix multiple, il doit choisir une réponse parmi celles qui lui sont proposées déjà élaborées.

Certains tests sont maintenant disponibles avec une possibilité de correction et même d'interprétation par ordinateur, d'autres devant être corrigés "à la main".

Les tests qui donnent un résultat chiffré débouchant sur un classement, et qui donc se réfèrent à un système de mesure, à une métrique sont qualifiés de "tests psychométriques". Les autres, qui proposent une interprétation directe des données sans le recours à la mesure, sont souvent regroupés sous le terme "d'épreuves cliniques".

GOGUELIN s'est essayé à dresser l'inventaire des tests en fonction du domaine qu'ils explorent. Il distingue douze catégories que nous évoquerons rapidement en citant quelques exemples.

1 - TESTS DYNAMO MUSCULAIRES

Ce sont des appareils qui enregistrent le travail du muscle dans différents aspects (ex: le test du dynamomètre à main).

2 - TESTS SENSORIELS

Ces tests mesurent des sensations telles...

- l'audition : mesure à l'aide de l'audiomètre;
- la vision : contrôle de l'astigmatisme avec le cadran de PARENT, contrôle de la perception des couleurs avec l'album de POLACK...
- les sensibilités proprioceptives : test de classement de poids de BINET...

3 - TESTS MOTEURS

Ils permettent l'étude de divers mouvements tels...

- mouvements répétitifs : "tapping", pointage...

- mouvement simples : pointillage (faire un point dans chaque case d'un quadrillage le plus rapidement possible)...
- dextérité manuelle : test des rondelles de PIORKOWSKI (enfiler des petites rondelles percées de 7 trous sur 4 tiges d'acier)...
- stabilité musculaire et nerveuse
- développement moteur: échelle d'OZERETSKY utilisable entre 4 et 16 ans...

4 - TEST SENSORI-MOTEURS

Il s'agit de l'étude des temps de réactions motrices à l'aide d'un chronoscope qui indique le temps écoulé entre le stimulus et la réponse.

5 - TESTS PSYCHOMOTEURS

Ils sont centrés sur l'étude des coordinations oculo-motrice, audio-motrice... (coordinations oeil et main, droite et gauche, membre supérieur et membre inférieur...) et d'une manière générale des capacités de contrôle, d'attention et de vigilance.

Le test du chariot tour de LAHY, par exemple, consiste pour le sujet à déplacer un stylet dans un chemin sans en toucher les bords et le plus rapidement possible à l'aide de deux manivelles.

BONNARDEL a mis au point plusieurs épreuves de ce type comme le "test oméga", le "test de la grecque", le "double labyrinthe" qui met en oeuvre une double coordination oculo-motrice dans la mesure où il faut contrôler simultanément la position de deux ergots dans deux chemins qui défilent à vitesse constante.

Le polyréactiographe est un appareil complexe qui regroupe des épreuves de temps de réaction et des épreuves psychomotrices...

6 - TESTS MENTAUX ANALYTIQUES

Ce sont ceux qui stimulent l'une des activités intellectuelles supérieures. On distingue :

- tests de facteur g (g comme général): épreuves qui s'appuient sur un seul type de tâche, la réussite à cette tâche étant représentative d'une intelligence générale, d'une capacité à réussir toutes sortes d'autres tâches. Le prototype en est le test des "progressive matrices" ou PM38 de RAVEN qui fait appel à l'éduction de relations (découverte de la relation qui unit deux éléments) et à l'éduction de corrélats (découverte d'un élément connaissant un premier élément et la relation qui unit ces deux éléments).

- tests composites d'intelligence: épreuves qui reprennent la conception de BINET selon laquelle la nature des tests importe peu pourvu qu'ils soient

nombreux et s'appuient sur l'idée d'une intelligence globale. On peut citer le "test army alpha" utilisé en 1914-1918 comme épreuve de sélection des soldats de l'armée américaine, le test de raisonnement de RENNES...

- activités inductives: l'induction est cette activité intellectuelle qui rapproche et compare un ensemble de faits de même nature pour en tirer une règle, un principe, une loi... Elle participe au développement psychologique de l'enfant, à la découverte, la formalisation et l'acquisition des connaissances. Le test d'éduction de PIRE met en relief cette importante capacité.

- formation de concepts: épreuve de HANFMANN et KASANIN, qui permet d'étudier la capacité à former des concepts.

- capacités métacognitives: mise en évidence de la capacité au sujet à prendre conscience de ses propres modes de fonctionnement cognitif et à les décrire. Voir le test d'évaluation des capacités métacognitives -T.E.C.M.- de PASQUIER.

- potentiel d'apprentissage: appréciation de la capacité d'apprentissage du sujet selon le principe d'une passation dynamique en trois phases. La première phase consiste à passer une épreuve. Au cours de la seconde un entraînement est proposé au sujet. La même épreuve est donnée après l'entraînement. L'évolution de la performance constitue un indice de la sensibilité du sujet à un apprentissage. L'entraînement peut se faire également au cours de la passation du test.

- tests de facteur verbal: étude du vocabulaire passif (test de vocabulaire de BINOIS PICHOT...) ou de la compréhension verbale par appariment de phrases de sens proche (test BV50 de BONNARDEL)...

- tests d'intelligence concrète: il s'agit d'épreuves qui font appel à la manipulation d'objets. Le prototype en est les cubes de KOHS, jeu de 16 cubes aux faces de différentes couleurs, identiques d'un cube à l'autre et avec lesquels le sujet doit construire des configurations de difficulté croissante et qui lui sont présentées sur une série de cartons. On peut citer aussi le "bloc de WIGGLY" parallélépipède rectangle de bois découpé en 9 pièces sinueuses à reconstituter le plus rapidement possible...

- tests d'attention: les classiques sont les "tests de barrage" initialiés par celui de BOURDON paru en 1895 où il fallait barrer certaines lettres dans un texte. Le test de TOULOUSE et PIERON (1904) et ses dérivés sont des plus employés: il s'agit de barrer trois des petits carrés dotés d'un trait extérieur adjacent disposé selon les huit orientations de la rose des vents dans une page de ces carrés disposés en lignes.

- tests de visualisation spatiale: ils explorent la capacité de se représenter dans l'espace une figure bi ou tri-dimensionnelle dessinée sur un plan.

- tests de célérité perceptive et mentale: c'est dans ce cas la rapidité de réponse qui est prise en compte ici.

- tests de capacité à exécuter des consignes.

- tests mécaniques: le test mécanique de RENNES propose la résolution de petits problèmes qui mettent en jeu des situations concrètes ne demandant pas de connaissances scolaires. Il est très utilisé pour la sélection des ouvriers spécialisés ou pour l'entrée en formation professionnelle...

- tests d'aptitude numérique: mesure de l'aptitude à manipuler rapidement les symboles, les nombres...

- tests de mémoire: étude du volume de la mémoire ainsi que de son mode d'organisation pour laquelle le test de la figure complexe de REY est un grand classique. Le sujet doit reproduire de mémoire une figure géométrique composée d'un rectangle "habillé" de nombreux sous-ensembles. On peut ainsi constater comment le sujet organise sa réponse, autour du rectangle de base, ou par le contour de l'ensemble, ou en juxtaposant les détails...

- tests de dépendance-indépendance à l'égard du champ: étude de la capacité de se détacher d'un champ visuel organisé pour en percevoir une partie séparée de l'ensemble. Par exemple, dans le "children's embedded figures test" de WITKIN, OLTMAN, RASKIN et KARP, il est demandé à l'enfant de retrouver un triangle, puis une maison, dissimulés dans le dessin d'un objet familier (on trouve ce genre de discrimination visuelle dans les albums pour enfants, quand il faut chercher et colorier un personnage fondu dans un arbre, dans un paysage).

- tests de créativité.

- tests d'aptitude à la musique.

- tests d'aptitude à l'esthétique...

7 - BATTERIES FACTORIELLES

On appelle batterie factorielle un ensemble de tests mentaux diversifiés susceptibles de recouvrir tout ou partie des capacités mentales isolées par les analyses en facteurs. Par exemple, la batterie de BONNARDEL destinée aux cadres débouche sur 6 notes distinctes: potentiel intellectuel, intelligence pratique, compréhension verbale, célérité perceptive et mentale, calcul, orthographe.

8 - TESTS MENTAUX SPECIFIQUES

Certains cas particuliers amènent l'élaboration et l'usage d'épreuves adaptées. Ainsi les tests pour les analphabètes, pour les travailleurs étrangers, pour les aveugles, pour les employés de bureau et autres professions...

9 - TESTS DE DEVELOPPEMENT

Cette famille de tests est utilisée lorsque le psychologue s'interroge sur le degré de développement de l'intelligence d'un individu. La performance du sujet est comparée aux performances moyennes des individus de différents groupes d'âges. On peut ainsi définir son "âge de développement" ou son "âge mental". Jusque vers 3 ou 4 ans on calcule un "quotient de développement", le Q.D., qui prend en compte essentiellement des indices psychomoteurs. Ensuite, jusque l'âge de 15 ans, on calcule un "quotient intellectuel", le fameux Q.I., selon la formule

$$Q.I. = \frac{\text{"âge mental"}}{\text{âge du sujet}} \times 100.$$

Dans d'autres cas, la performance du sujet est simplement comparée aux performances des sujets du même âge.

Pour les âges adultes, on prend en considération des indices statistiques de dégradation des capacités intellectuelles.

- tests de développement pour la première enfance: on peut citer l'échelle de développement psychologique de la première enfance de BRUNET-LEZINE qui couvre les âges de 1 mois à 6 ans.

- tests de quotient intellectuel: le plus célèbre, qui a inspiré tous les autres, est l'échelle métrique de l'intelligence crée en 1905 par BINET et SIMON (ces auteurs n'utilisaient pas le Q.I. proposé ultérieurement par STERN, mais exprimaient le résultat en termes de mois de retard ou d'avance dans le développement). La dernière adaptation, la nouvelle échelle métrique de l'intelligence, -la N.E.M.I.-, date de 1949 et est l'oeuvre de ZAZZO, GILLY et VERBA.

Théoriquement, il est possible de construire un test de Q.I. en utilisant toute tâche dont la réussite moyenne évolue avec l'âge. Ainsi, les labyrinthes de PORTEUS donne une idée de l'intelligence dans sa dimension adaptabilité sociale. Le dessin du bonhomme a été coté pour donner un âge mental par GOODENOUGH.

Les cubes de KOHS et ses multiples adaptations permettent de calculer des Q.I. LONGEOT s'est inspiré des travaux de PIAGET pour concevoir une échelle de développement de la pensée logique.

Dans le W.I.S.C., WECHSLER intelligence scale for children, l'auteur a séparé une partie "verbale" d'une partie "performance". Des épreuves à passation collective ont été élaborées par GILLE, par BOURDIER... pour un dépistage en milieu scolaire...

- pour les adultes: l'épreuve de WECHSLER pour adultes -W.A.I.S.- est la plus utilisée.

10 - TESTS DE CONNAISSANCES

Ces tests évaluent les connaissances des sujets. On peut envisager la construction de tests relativement à tous les domaines de la connaissance. Pratiquement, ce type de tests est largement utilisé en psychologie scolaire, en orientation scolaire et professionnelle, en formation professionnelle des adultes, en psychologie du travail, en psychopédagogie... Un chapitre suivant est consacré à cette famille de tests.

11 - TESTS DE MOTIVATION

Les épreuves regroupées dans cette rubrique explorent l'existence et l'intensité d'un intérêt particulier pour tel ou tel domaine d'activité scolaire ou professionnelle. Ces épreuves sont utilisées pour le conseil en orientation, pour le processus de sélection...

-I.R.M.- de ROTTWELL-MILLER permet de cerner la motivation professionnelle d'une personne à travers 12 catégories: plein air, mécanique, calcul, science, contacts personnels, esthétique, littérature, musique, services sociaux, travaux de bureau, activités pratiques, activités médicales. Le test d'intérêt pour les situations scolaires -T.I.S.S.- de PASQUIER facilite l'expression des attitudes développées par les jeunes face à l'école. Les questionnaires de motivation -FIVA2- de satisfaction -FIVA1- de FRANCES sont préconisés dans l'entreprise comme outils de diagnostic du dysfonctionnement social manifesté par l'absentéisme, les démissions...

12 - TESTS DE TEMPERAMENT, DE CARACTERE ET DE PERSONNALITE

Cette douzième et dernière catégorie comprend toutes les épreuves qu'on ne peut pas classer dans les onze catégories précédentes. Définie positivement, on retrouve ici l'étude des traits de personnalité, soit d'une manière analytique, centrée sur un trait ou une dimension déterminé, ou sur un ensemble de traits définissant un profil, soit d'une manière globale, synthétique centrée sur la définition de la personne. On note deux sous-groupes: les questionnaires qui, comme leur nom l'indique, proposent au sujet une liste de questions auxquelles il est demandé de répondre par oui ou par non, parfois de placer sa réponse sur une échelle bipolaire et les "tests projectifs" qui proposent diverses situations à résoudre.

Parmi les questionnaires, on peut citer l'échelle d'auto-estimation de LAVOEGIE basée sur neuf types de comportement: énergie-dynamisme, sociabilité, ténacité, intelligence sociale, ambition-confiance en soi, réflexion-circonspection, optimisme, sincérité, pondération. L'inventaire de tempérament de GUILFORD-ZIMMERMAN conduit à distinguer dix facteurs: activité générale, contrainte, ascendance, extraversion ou introversion sociale, stabilité émotionnelle, objectivité, bienveillance, tendance à la réflexion, relations personnelles, masculinité. Le plus élaboré des questionnaires qui donne des typologies dans les termes d'une psychologie commune est le 16PF de CATTEL et EBER. Les 16 facteurs de personnalité isolés sont les suivants: A sizothymie-affectothymie (réservé-ouvert); B intelligence; C force du moi (stabilité

émotionnelle); D ascendance – soumission (caractère affirmé – humble); E expansivité; G force du surmoi (consciencieux – insouciant); H threctia – parmia (timide – aventureux); I harria – premsia (dur – tendre); L alaxia – protension (confiant – méfiant); M praxernia – autia (pratique – imaginatif); N naïveté – clairvoyance (direct – fin); O calme – anxieux; Q1 conservatisme – radicalisme; Q2 dépendant du groupe – se suffisant à lui – même; Q3 intégration (incontrôle – contrôlé); Q4 tension ergétique (détendu – tenu). En seconde analyse, on peut situer le sujet sur quatre dimensions: anxiété, extraversion, sensibilité, indépendance.

Quelques questionnaires, tel l'inventaire multiphasique de personnalité du Minnesota, s'inspirent de catégories nosologiques utilisées par la psychiatrie: hypocondrie, dépression, hystérie, personnalité psychopathique, masculinité – féminité, paranoïa, psychasténie, schizophrénie, hypomanie, introversion sociale.

Les tests projectifs proposent au sujet une situation souvent peu structurée dans laquelle la personnalité du sujet est incitée à se projeter en interprétant une réalité ambiguë. Le matériel proposé peut être totalement informel comme dans le test psychodiagnostic de RORSCHACH qui est basé sur l'interprétation de tâches d'encre. Il peut aussi n'y avoir aucun matériel quand on demande à un enfant de dessiner une famille (épreuve de CORMAN), ou à un adulte de dessiner un arbre (épreuve de KOCH).

Le matériel peut inviter à la combinaison créatrice comme dans le test du village imaginaire (LOWENFELD, ARTHUS, MABILLE, MUCCHIELLI). Enfin, les épreuves dérivées du test de VAN LENNEP, telles le "thematic apperception test" – T.A.T. – de MURRAY et BELLAK, le test dynamique personnelle et images – D.P.I. – de PERRON, le test de frustation de ROSENZWEIG, le test "patte noire" de CORMAN... proposent d'imaginer et de raconter des histoires à partir d'une scène ou d'une suite de scènes présentées sous la forme d'images.

Une catégorisation reprise par LEON, plus simple et en trois groupes est de plus en plus utilisée. On distingue en premier lieu les tests de connaissances qui contrôlent un savoir quelconque et sur lesquels nous reviendrons longuement. Les tests relevant du domaine cognitif, on si l'on veut dire vite les "tests d'intelligence", derrière leur grande diversité font tous appel à des raisonnements intellectuels. Enfin, les tests concernant des aspects non cognitifs de la conduite, ou pour faire plus court les "tests de personnalité", explorent les différentes facettes psychologiques du sujet.

Suite à ce survol panoramique, on peut donc dire des tests, d'une manière très générale, qu'ils sont des instruments d'observation et de recueil de données. Toutefois, ils se distinguent par quelques caractéristiques précises des autres techniques de recueil et de traitement de l'information.

La situation dans laquelle on place le sujet, conçue pour stimuler tel ou tel comportement, ou pour provoquer telle ou telle réponse, est standardisée. Cela veut dire que tous les paramètres qui participent à la construction de la situation de test sont rigoureusement définis à l'avance et une fois pour toutes. A priori, et dans l'idéal, l'utilisation d'un même test à différents moments et dans différents lieux reproduit exactement la même situation: même disposition topologique entre

testeur et testés, **même matériel** utilisé, **mêmes consignes verbales, même attitude du** testeur et même type de réponse aux questions éventuelles des testés...

Le testeur adopte une **attitude de neutralité bienveillante.** D'un côté il encourage le sujet à donner le meilleur de lui-même et de l'autre il s'interdit de donner tout indice qui puisse l'aider à trouver la réponse attendue par crainte de biaiser le résultat: ce sont les connaissances ou les capacités du sujet qui sont l'objet de l'évaluation et non celles de l'examinateur.

La construction du test permet une **notation du comportement ou de la réponse sans ambiguïté.** Dans l'idéal, deux examinateurs travaillant en parallèle sur la même passation d'un test devraient arriver à des conclusions identiques. Cela implique qu'ils focalisent leur attention sur les mêmes éléments de la réponse, qu'ils appliquent le système de cotation préconisé et que ce système prévoit toutes les possibilités de répondre accompagnées de leur mode de notation ou de catégorisation.

Le test implique la **comparaison statistique de la performance du sujet avec celles obtenues par les sujets d'un groupe de référence.** Cette comparaison débouche sur le classement du sujet soit dans une catégorie particulière (attribution d'un certain type de personnalité par exemple), soit sur un continuum (le sujet possède plus ou moins un trait psychologique), soit sur une échelle numérique (le sujet a un Q.I. de 115 points).

On comprend maintenant plus facilement les nécessités de la standardisation de la situation et de la correction, de l'attitude neutre de l'examinateur comme conditions d'obtention d'un classement juste.

Notons qu'il arrive de rencontrer le mot test utilisé dans un sens étendu à toute situation de contrôle: le test d'anglais qui figure dans le livre de classe n'a que peu à voir avec le test psychométrique accompagné d'un dispositif statistique. Le "faux test" propose seulement une notation, au mieux une évaluation analytique par objectifs ou points de maîtrise. Le test digne de ce nom met à disposition une procédure de mesure élaborée selon des principes scientifiques et techniques rigoureux présentés au chapitre suivant.

Avant d'aller vers les aspects techniques des tests, et pour rester sur le plan des généralités, nous évoquerons rapidement leurs principaux domaines d'application. Les tests accompagnent le psychologue dans la diversité de ses activtés. Le chercheur en psychologie va utiliser les tests pour constituer différents groupes de sujets. Il peut les inclure dans son dispositif expérimental comme outil de mesure de la variable dépendante... Le psychologue du travail utilise les tests pour sélectionner des candidats à un poste ou à une fonction dont il a analysé les caractéristiques au préalable. Il peut recourir à leur usage comme outil de diagnostic d'un dysfonctionnement social ou technique... Le spécialiste de la psychopathologie peut évaluer le vieillissement différentiel des capacités, aborder l'étude des troubles d'une personnalité, dessiner la courbe évolutive des effets d'un traitement... en intégrant le test dans sa démarche clinique. Le psychologue de l'éducation aborde les difficultés scolaires individuelles dans le cadre de l'examen psychologique de l'écolier mené à l'aide de tests de connaissances pour apprécier le niveau scolaire,

à l'aide de tests d'intelligence pour mesurer l'efficacité des capacités cognitives, à l'aide d'épreuves de personnalité pour dégager les éléments psychologiques d'une éventuelle perturbation mentale...

Quelques échos dans les médias ont semblé donner une certaine toute puissance aux tests. C'est une opinion qui n'est pas justifiée. Le test est un outil, avec son mode d'emploi, avec les services qu'il peut rendre, avec ses qualités et ses défauts, avec ses limites qu'il faut bien connaître. Ce n'est pas tant le test qui est en cause, il est ce qu'il est, que l'usage ou plutôt les abus d'usage qui ont pu être relevés ici ou là. Il faut bien savoir qu'un test psychométrique donne un résultat chiffré qui permet de classer un individu par rapport à un groupe de référence et d'un point de vue défini par son contenu (par exemple, classement de la performance en anglais écrit, classement de la performance sur une échelle de développement...). Le chiffre obtenu est un événement objectif indiscutable en tant que fait daté et contextualisé. Le problème commence avec l'interprétation du fait et les questions posées à ce moment-là sont des questions délicates qui nécessitent prudence et connaissance. Le résultat obtenu au test donne-t-il la mesure du maxima que l'individu peut produire, est-il le reflet de sa capacité réelle, est-il l'expression fidèle de son potentiel? On n'en sait rien et le test ne nous renseigne pas. La capacité mesurée va-t-elle rester au niveau actuel ou bien va-t-elle évoluer? On n'en sait rien et le test ne nous renseigne pas mieux. Conclure hâtivement sur ces questions amène des critiques, voire des résistances justifiées. Il faut considérer que la décision qui peut être prise sur la base d'un résultat au test engage parfois une trajectoire sociale pour de nombreuses années: l'admission dans l'enseignement spécial n'est pas une garantie de réinsertion dans les classes normales; une orientation scolaire puis professionnelle définit l'activité future; le refus d'une embauche peut amener dépressions et suicides; l'expertise psychologique judiciaire influence la décision d'un jury...

Il faut donc laisser le test à sa place et ne pas essayer de lui faire dire ce qu'il ne peut pas dire. De fait, le test est souvent à considérer comme l'amorce d'un questionnement. Si un élève obtient un mauvais classement à un test de mathématiques, cela doit stimuler une analyse du problème posé. Tout d'abord, il convient de se centrer sur les causes possibles de la mauvaise performance. En premier lieu, il faut s'assurer de la manière de fonctionner de l'élève afin de mettre en évidence le processus qui a amené la réponse. Pour cela on peut lui faire recommencer quelques questions en lui demandant de raisonner à haute voix par exemple. On peut alors commencer à cerner la nature de l'échec: incompréhension globale, non perception du problème, raisonnement partiel, raisonnement correct s'appuyant sur des données fausses, erreurs dans les calculs numériques, lacunes notionnelles plus ou moins étendues... A partir de là, on peut formuler des hypothèses qui vont guider le choix vers de nouveaux tests ou autres techniques d'examen. Il faut aussi s'intéresser aux données contextuelles, voir les résultats habituellement obtenus en mathématiques, -nul n'est à l'abri d'un "accident de parcours"-, voir la résistance au stress des situations d'examen, discuter des conditions particulières au jour de l'examen (état de fatigue ou problème de santé, mauvaises relations sociales intrafamiliales passagères ou continues, mauvaises relations sociales intrascolaires avec les autres élèves, avec les maîtres, mauvais résultat obtenu dans une autre matière), discuter avec l'élève de son goût pour la discipline, de son investissement et de ses efforts,

du sentiment de sa compétence dans cette matière, observer les aspects culturels du test qui font que le contenu en sera plus ou moins familier... Et à chaque fois que subsiste un doute, on peut donner une forme parallèle du test ou un autre test qui mesure le même objet.

Les tests sont donc utiles pour établir des constats de la manière la plus objective qui soit. Ce qui est fait de ces constats relève non pas du test mais de l'examinateur et de son art de l'interprétation.

CHAPITRE 10

COMMENT CONSTRUIRE UN TEST ?

Après avoir considéré les tests d'une certaine distance au chapitre précédent nous proposons au lecteur de rentrer à l'intérieur d'un test en suivant les étapes de sa construction. De la sorte, le lecteur pourra élaborer une représentation détaillée de l'objet test et juger en connaissance de cause de l'intérêt ou du non-intérêt de l'utilisation des épreuves standardisées, des **tests pédagométriques** dans l'évaluation scolaire et/ou formative. Imaginons un instant qu'un Martien arrivant de sa planète et débarquant sur la Terre pour la première fois y découvre une calculatrice de poche. Ce n'est pas parce qu'il va la démonter en ses différents pièces et composants qu'il parviendra à comprendre son fonctionnement et à l'utiliser efficacement! C'est en essayant, par la pratique, qu'il pourra progressivement acquérir un usage efficace de cet outil. Il ira d'aurant plus vite dans son apprentissage qu'un pédagogue-médiateur s'interpose entre lui et cet apprentissage pour lui fournir les aides les plus efficaces. C'est pourquoi nous encouragerons vivement le lecteur intéressé à essayer en grandeur nature la construction didactique du test qui est décrite ci-après.

Nous proposons un contenu et une tâche simples afin de pouvoir être entièrement disponible pour se centrer sur les aspects techniques. Le problème posé et son traitement préconisé pourront sembler bien simplistes et artificiels pour les professionnels de la psychologie et de la psychométrie. Mais ce chapitre destiné à l'information des pédagogues et des formateurs a été conçu dans une forme didactique qui réponde à un premier niveau d'information. Les variations techniques, la justification des lois mathématiques utilisées, les débats qui agitent le "microcosme" des psychologues sur telle ou telle subtilité technique ne seront pas discutés dans le cadre de cet ouvrage.

Le contexte est le suivant: une entreprise de travaux de secrétariat et de traitement de textes veut augmenter son équipe d'opératrices qui a pour tâche la préparation d'envois postaux en nombre. Les adresses sont recopiées à la main à partir de listes. L'entreprise passe donc un contrat avec un établissement spécialisé dans la psychologie du travail pour le recrutement d'une dizaine d'employées.

Le psychologue chargé du traitement de cette commande décide de construire une épreuve qu'il va inclure dans le processus de recrutement qu'il pense mettre en oeuvre. En premier lieu, il analyse les paramètres qui interviennent dans un travail de copie d'adresses. Cette tâche présente un caractère répétitif. Elle demande une grande rapidité d'exécution sur le plan psychomoteur. Dans l'aspect répétitif s'introduisent toutefois des éléments de variation. Les caractères peuvent être minuscules ou majuscules. On copie des lettres et des chiffres (numéros de rue, codes postaux). L'aspect purement moteur de la tâche est pondéré par des facteurs cognitifs tels un minimum de connaissances en lecture -plus le mécanisme de lecture est automatisé et fluide et plus le travail sera fait rapidement- ou encore en géographie -le degré de familiarité avec les noms à lire puis à copier va diminuer le temps de déchiffrage-. Ajoutons à cela que le temps disponible pour assurer l'ensemble des opérations de recrutement devra être limité du fait que le négociateur du contrat a dû consentir à un "petit prix" en conséquence d'une vive concurrence d'une part et pour fidéliser le client d'autre part. Le test doit donc prendre un temps minima de passation et de correction.

Finalement, le psychologue élabore l'épreuve suivante: les candidats devront se livrer à quatre exercices différents.

- Premier exercice: copie de la lettre "a". Cet exercice touche uniquement à la rapidité gestuelle.

- Deuxième exercice: copie de la cellule graphique "a A". Une rupture est introduite dans la mesure où la cellule à reproduire contient l'association de la minuscule et de la majuscule. L'aspect rythme prend ici de l'importance.

- Troisième exercice: copie de la suite "a1 b2 c3..." Ici interviennent deux nouveautés. Chaque lettre est associée à un chiffre. La suite des lettres remplace le simple "a". La combinaison de ces deux nouveautés fait que chaque lettre de l'alphabet est accompagnée de son rang. On a donc l'introduction d'un facteur cognitif.

- Quatrième exercice: copie de la suite "aA1 bB2 cC2..." On combine les deux associations précédentes ce qui fait que chaque lettre dans l'ordre de l'alphabet est associée de sa majuscule et de son rang. A la rapidité, au rythme, à la cognition s'ajoute la flexibilité mentale puisqu'il faut changer très souvent de catégories, -minuscules, majuscules, nombres entiers-, tout en respectant les relations d'ordre intra-catégorielles et les relations inter-catégorielles.

On peut conclure que le test recouvre bien a priori un ensemble de capacités fonctionnelles nécessaires à la copie rapide.

Pour chaque exercice les sujets disposeront d'une minute. Entre chaque exercice une minute de récupération est accordée. Le test se fait sur papier libre et à l'aide d'un stylo feutre noir à pointe fine. Voilà pour le choix de la tâche, le rationnel et la structuration des exercices et du matériel. Il reste à écrire le texte de la consigne.

La consigne donnée doit être interprétable de manière univoque pour tous les sujets. Il faut en conséquence la rédiger dans un langage simple mais non simpliste, en choisissant les mots et les tournures syntaxiques qui aient la probalité la plus élèvée d'être compris de toutes les personnes qui auront à passer le test. D'autre part cette compréhension doit être la même pour tous. En conséquence, il convient d'exclure tous les mots polysémiques, tous les mots aux sens vagues ou ambigus, tous les mots appartenant à des jargons, tous les idiotismes... Au travers du sens des mots, la lecture ou l'écoute de la consigne doit mettre tous les sujets dans la même disposition à effectuer le même type de tâche. La réalisation de cette tâche est supposée mettre en évidence les capacités que le test doit mesurer.
(Note: l'emploi de l'italique signale les propos d'ordre général; l'écriture droite se rapporte à la construction du test de rapidité graphique).

Par exemple, la consigne "vous recopiez ce que j'écris au tableau" est incomplète et risque de ce fait d'induire des comportements différents d'un individu à l'autre et en fin du compte non comparables. A aura écrit sur une feuille 21 X 27, B sur un tout petit carnet, C sur une ardoise, D avec une machine à écrire,... A a pris le temps de calligraphier sa copie et de l'ornementer à la manière des moines copistes, B a écrit à toute vitesse et de son écriture personnelle illisible par autrui, C qui est immigré a copié le texte dans sa langue d'origine, D qui a un travail urgent à terminer a décidé qu'il recopiera le texte demain... Les résultats obtenus ainsi, du fait du caractère implicite de la consigne, ne pourront pas être comparés et se révèleront inexploitables en termes d'évaluation. Dans le cas présent, ce sera l'idée de rapidité qui sera privilégiée, à un niveau minimum de lisibilité. Le texte suivant est finalement retenu :

"Vous allez subir une série de 4 épreuves de rapidité d'écriture.

Au signal donné, vous allez reproduire sur une feuille et le plus rapidement possible le modèle que je ferai au tableau. Attention, votre écriture doit rester lisible.

Vous stopperez immédiatement au second signal.

Pour commencer, vous reproduirez la lettre a minuscule.
(écrire le modèle au tableau "a a a ...")

Attention... Partez !
(déclencher le chronomètre, laisser travailler 1 minute précise)

Stop! Vous avez droit à une minute de repos.
(au bout d'une minute passer à l'exercice suivant)

Pour le second exercice, vous reproduirez le plus rapidement possible la lettre a minuscule suivie de la lettre a majuscule.
(écrire le modèle au tableau "aA aA aA ...")

Attention... Partez !
(déclencher le chronomètre, laisser travailler 1 minute précise)

Stop! Vous avez droit à une minute de repos.
(au bout d'une minute passer à l'exercice suivant)

Pour le troisième exercice vous, reproduirez le plus rapidement possible chaque lettre de l'alphabet suivie du nombre qui indique son rang. Si vous arrivez à la lettre z avant le signal d'arrêt vous recommencez une nouvelle série à partir de a.
(écrire le modèle au tableau "a1 b2 c3 ...")

Attention... Partez !
(déclencher le chronomètre, laisser travailler 1 minute précise)

Stop! Vous avez droit à une minute de repos.
(au bout d'une minute passer à l'exercice suivant)

Pour finir, le dernier exercice consiste à reproduire le plus vite possible dans le temps donné la suite des lettres de l'alphabet, chaque lettre minuscule étant suivie de sa majuscule et du nombre qui indique son rang.
(écrire le modèle au tableau "aA1 bB2 cC3 ...")

Attention... Partez !
(déclencher le chronomètre, laisser travailler 1 minute précise)

Stop! C'est terminé, je vous remercie de votre participation".

Dans le cours d'une passation de test, avant de lancer la consigne, il est important de préparer les aspects matériels. Puisque notre épreuve vise à mesure la rapidité d'écriture des sujets, la qualité de l'installation physique du sujet ne doit pas être négligée. Les tables seront individuelles, les sièges à hauteur règlable, la surface du plateau sera suffisamment lisse sinon il faut prévoir la fourniture d'un sous-main. La température ambiante de la salle et sa ventilation, le degré du bruit produit par l'environnement, la qualité de l'éclairage, la disposition du tableau et sa lisibilité de tous les points de la salle... sont à vérifier.

Le matériel mérite également une vérification attentive.

Donner un stylo qui fonctionne mal ou pas du tout, ou bien manquer de papier dans une épreuve de rapidité d'écriture gâchent définitivement l'utilisation du test: il faut relancer l'exercice pour tout le groupe et les résultats n'auront pas été obtenus dans des conditions standardisées. D'autre part, "l'image de marque" du psychologue ira se dégradant, dans le sens d'un manque de sérieux.

Bien entendu on ne lance pas une consigne de test de manière abrupte! Il est préférable de développer un préambule qui puisse rassurer les sujets, les détendre – en démythifiant ce qu'est un test par exemple – tout en instaurant un climat propice à la concentration et aux investissements affectif et énergétique dans la tâche.

Une fois l'élaboration du test terminée, l'étape suivante consiste à construire un ***étalonnage:*** *"On étalonne un* ***test*** *en l'appliquant à un groupe de référence, à un groupe étalon, ce qui permet d'établir des normes dans l'emploi du* ***test étalonné"*** *(PIERON).*

Les phases de l'étalonnage sont les suivantes :

1 – constitution du groupe de référence,
2 – passation de l'épreuve,
3 – correction et notation des épreuves,
4 – mise en forme des données numériques,
5 – construction de l'étalonnage.

1 – Le groupe de référence doit être représentatif de la ***population parente à*** *laquelle le test va s'adresser. Si on travaille un test de français pour le cours élémentaire 2me année, le groupe de référence qui va permettre d'étalonner le test devra donner une image fidèle de tous les élèves de tous les C.E.2 de France et de Navarre. Cela revient à former un sous – ensemble qui représente au mieux la population virtuellement intéressée par la passation du test. Les instituts de sondage connaissent le même type de problème. Ils n'ont pas le moyen, ni le temps, ni même la possibilité d'interroger tous les français et toutes les françaises sur tel ou tel point. En général, ils se contentent de s'adresser à un groupe réduit d'environ mille personnes. Mais cet "échantillon" est censé représenter les opinions de l'ensemble de la population. Pour le test, c'est pareil. Comment constituer un groupe de référence valable, qui représente bien ce qu'il doit représenter ?*

Trois méthodes sont à considérer. ***L'échantillonnage au hasard*** *par lequel on fait confiance au hasard pour répartir les caractéristiques de la population selon une même proportion dans l'échantillon. Ainsi, pour notre test de français au C.E.2., on tire mille élèves de ce cours au hasard. Si le tirage au sort est réalisé correctement, chaque personne de la population à la même chance de participer au groupe étalon.*

Pour notre test de rapidité d'écriture, on fait passer une annonce d'offre d'emploi dans la presse et on donne l'épreuve à toutes les personnes qui vont y répondre favorablement.

La seconde méthode est ***l'échantillonnage stratifié*** *qui va prendre en compte certaines des catégories de la population. Chaque catégorie retenue devient une strate dans laquelle les individus seront tirés au sort pour former le groupe de référence. Ainsi, pour le test de français du C.E.2 on peut considérer et alimenter deux strates garçons et filles. On peut en considérer neuf en reprenant les catégories socio – professionnelles de l'I.N.S.E.E.. On peut en considérer dix – huit en*

combinant les deux critères de sexe et d'origine socio-professionnelle. Mais le niveau scolaire général est aussi un facteur important pour un test de français, ainsi que l'âge et l'origine ethnique. Et puis certains élèves sont des redoublants, certains lisent beaucoup à la maison et d'autres pas du tout... Tenir compte de toutes ces catégories en les combinant constitue vite un casse-tête insurmontable. Les strates les plus usuelles restent l'âge, le sexe et le milieu.

Pour notre test d'écriture, les catégories intéressantes seraient l'âge pour tenir compte de l'effet du vieillissement biologique sur la rapidité psycho-motrice et l'expérience professionnelle en rapport avec l'écrit pour tenir compte du degré de familiarité avec ce type de travail.

La troisième méthode est une variante de la seconde. Elle est dite ***échantillonnage par quota*** *ou* ***échantillonnage proportionnel*** *ou* ***échantillonnage stratifié pondéré.*** *On conserve le choix de catégories pertinentes et en plus on va alimenter les strates de l'échantillon dans les mêmes proportions que dans la population représentée. Par exemple, s'il y a 8% de redoublants au C.E.2 en moyenne et sur le plan national, la strate redoublants de l'échantillon comportera 8% de son effectif total.*

Pour notre étude imaginaire nous retenons le principe de l'annonce comme le moyen le plus économique, le plus rapide, le mieux adapté pour former un échantillon représentatif des personnes susceptibles de se porter volontaires à l'embauche pour ce travail. Admettons que suite à la diffusion de cette annonce dans la presse une trentaine de candidates se soit faite connaître.

2 – Cela permet d'envisager l'étape suivante: la passation de l'épreuve. *On sait que la passation peut être soit individuelle, soit collective. La passation individuelle permet à l'examinateur d'établir un contact direct avec le sujet. Elle facilite l'observation des comportements et le recueil de données qualitives sur le mode de fonctionnement du sujet, sur ses stratégies de résolution, sur les types de raisonnement qu'il met en oeuvre, sur son niveau de langage et sur son expression spontanée, sur son style de réaction psycho-affective face à la difficulté ou à la monotonie, sur sa capacité à ouvrir et alimenter une relation sociale... Toutes ces données observées vont alimenter le commentaire qui accompagnera la mesure de la performance. De plus elles sont une source d'information précieuse pour une connaissance plus fine de l'individu. Le commentaire qualitatif participe à la compréhension globale de l'individu et aussi à la compréhension dans le détail d'un comportement spécifique. Prenons un exemple de test individuel de lecture, celui de BOREL-MAISONNY, qui a pour objectif de mettre en évidence les difficultés de prononciation éventuelles de certains sons. Il est demandé à l'enfant de lire des listes de "mots" sans signification comme celles-ci :*

mouko	*vafitaruder*
bartin	*sanzibidélu*
fanvé	*pudounirital*
yèroi	*munignaméso*
linou	*pulblagoritel*
chanedu	*zoltiduseltor*

gontra	*anslingélitil*
zulseu	*varduostivar*
lurir	
panbi	

Il est bien évident qu'une lecture collective de ces listes n'aurait aucun sens. La passation individuelle, au contraire, donnera comme résultat la liste des sons que l'enfant a du mal à prononcer, soit qu'il ne le sache pas du tout, soit qu'il prononce partiellement, soit qu'il déforme le son... D'autres observations seront possibles comme l'éventualité de l'apparition d'un bégaiement, d'une expression émotive excessivement tournée vers l'inhibition ou vers l'agitation. La réaction vis à vis d'un matériel phonologique sans signification pour l'enfant est intéressante à noter. De la synthèse de cet ensemble pourront sortir des indications précises utiles à la rééducation orthophonique si elle s'avère nécessaire.

On voit bien à travers cet exemple que la passation individuelle s'impose à chaque fois qu'il y a nécessité d'une connaissance fine d'un comportement, d'un fonctionnement, d'un mode de réaction... Elle est également indispensable avec des sujets non-lecteurs qui ne peuvent ni comprendre des consignes écrites, ni produire leurs réponses par écrit. Dans les faits de la psychologie de l'éducation par exemple les modes de passation collectif et individuel seront combinés: de la passation d'une épreuve collective, appelée parfois "épreuve de dépistage", on retire la liste des élèves ou des stagiaires qui ont des niveaux de performance au test jugés insuffisants. Ces personnes sont ensuite vues individuellement afin de mieux comprendre les sources de leurs difficultés et de rechercher le meilleur moyen de surmonter les-dites difficultés.

Dans notre petite histoire, dans la mesure où le but est simplement de construire un étalonnage des performances en rapidité d'écriture, la passation sera collective. Elle suivra les principes évoqués plus haut: préparation de la salle et du matériel, accueil et mise en confiance des candidats, suivi rigoureux de la consigne. A la fin de l'épreuve, il convient d'informer les candidates de la suite qui sera donnée: information par courrier par exemple. Suite à la passation, on se retrouve face à un matériel qu'il faut corriger et noter.

3 - La correction consiste à différencier bonnes et mauvaises réponses et la notation à attribuer des points aux bonnes réponses, éventuellement à en retirer pour les mauvaises.

Toute la question de la correction-notation se résume ainsi: les réponses, d'un même sujet corrigées par plusieurs correcteurs doivent toujours donner une note identique. C'est facile à dire mais difficile à obtenir! Vous demandez à un enfant de dessiner un carré. S'il dessine un triangle, la réponse est fausse. Mais à partir de quel écart de longueur dira-t-on que c'est un rectangle, à partir de quel écart de longueur et d'angle dira-t-on que c'est un quadrilatère quelconque? La précision de ces critères est nécessaire à la standardisation de la correction même si parfois elle prend un caractère arbitraire. L'essentiel est l'identité de traitement des réponses et en conséquence tous les cas de figure devront être prévus à l'avance.

Que décider pour notre test-exemple? Le test évalue la rapidité d'écriture, c'est-à-dire une capacité psycho-motrice. Il a été demandé que les lettres restent lisibles. L'appréciation de cette lisibilité se révèle a priori délicate. Bien sûr il ne s'agit pas d'évaluer la calligraphie mais de décider si la lettre est reconnaissable ou non. Après avoir défini le niveau d'exigence, on peut choisir comme règle que les réponses seront jugées fausses dans deux circonstances: le graphisme ne représente rien (c'est un gribouillis quelconque); le graphisme représente une autre lettre (au lieu d'un a on voit un o, un u, un c et un i). A partir du second exercice, on demande une association de caractères. Que faire si cette association n'est pas respectée? Deux cas peuvent se présenter. Soit la consigne n'a pas été comprise et de bout en bout l'association n'est pas respectée, soit on note des erreurs de temps à autre. Dans le premier cas, on peut dire que l'exercice n'est pas recevable. En effet, le sujet qui a mal interprété la consigne s'est situé différemment des autres sujets et on ne peut intégrer sa performance dans l'étalonnage (on dit que c'est un sujet "déviant" par rapport au test). Dans le second cas, il faut décider si une erreur d'association pénalise tout ou partie de la cellule à reproduire. Par exemple, dans le quatrième exercice, on relève une erreur dans la cellule "dD3" puisque la lettre d est au quatrième rang dans l'alphabet et non au troisième. Doit-on considérer que l'ensemble de la cellule est irrecevable ou bien prendre en compte qu'elle est correcte aux 2/3? Considérant que le but est de recruter des personnes qui vont recopier des adresses et qu'elles ne devront pas commettre d'erreurs si l'on veut que le courrier arrive à destination, l'option "tout ou rien" sera retenue et la cellule "dD3" ne fera marquer aucun point. Il peut arriver que le décalage numérique se poursuivre un certain nombre de cellules: "eE4 fF5 gG6 hH7 iI8 jJ10". La persistance du décalage marque un manque de contrôle inter-catégories puisqu'il n'y a pas mise en correspondance exacte du rang alphabétique et de la lettre. Il est décidé dans ce cas un compromis entre la rapidité et l'exactitude de l'exécution: la première cellule erronée est notée 0; pour les suivantes, les lettres lisibles sont notées 1/2 point, le total étant arrondi à l'unité supérieure. Le même principe est appliqué pour un décalage de lettres. Deux décalages corrélatifs amènent à compter faux. (On voit que même pour un simple test psycho-moteur les cas de figures à envisager au préalable peuvent être nombreux).

Une fois les principes de la correction définis, il faut établir un barème de notation, c'est-à-dire les règles d'attribution des points aux réponses. Le système le plus simple est de mettre un point par réponse juste. Dans certains cas, on observe que les erreurs non seulement ne rapportent rien mais sont sanctionnées par une pénalité, par des points en moins. D'autres fois un système de pondération est utilisé qui fait en sorte que soit prise en compte la difficulté relative de chaque question. Ainsi, certaines réponses seront notées 1 point, d'autres plus difficiles 2, 3 ou 4 points. En toute rigueur, le système de pondération devrait s'appuyer sur des fréquences de réussite réellement observées.

Pour notre test, nous choisissons la solution la plus simple: 1 point par lettre ou chiffre correct, à la condition applicable à partir du deuxième exercice que la cellule soit juste. Munis de nos critères de correction et de notre barême de notation, nous pouvons maintenant opérer sur les épreuves élaborées par les candidates puis passer à l'étape suivante.

4 – Mise en forme des données numériques. Suite à la notation des 30 épreuves, on se retrouve devant une série de 30 notes que voici :

294 291 280 295 286 301 293 297 302 291 305 295 298 308 294
297 300 304 301 305 293 311 287 300 285 305 287 299 297 295

Ces nombres représentent les **notes brutes** obtenues au test. (Cette série est construite artificiellement pour "les besoins de la cause". Nous encourageons le lecteur intéressé à faire l'expérience en grandeur nature et ensuite à traiter les données numériques qu'il aura recueillies. Il est également possible de traiter d'autres séries de notes recueillies à d'autres occassions, avec d'autres supports d'exercices...)

Une liste de notes ne parle guère. En première approximation elle permet de dire si les notes sont proches l'une de l'autre ou éloignées: en comparant la plus petite et la grande on a une première idée de ***l'étendue de la dispertion des notes.*** Elle donne le matériel nécessaire à la construction de l'étalonnage. Ce matériel va être transformé pour faire ressortir la forme des résultats, ***la forme de la distribution.*** On distingue classiquement différentes formes de distributions. La plus célèbre est ***la courbe de Gauss,*** dite courbe "en cloche" (fig. 1).

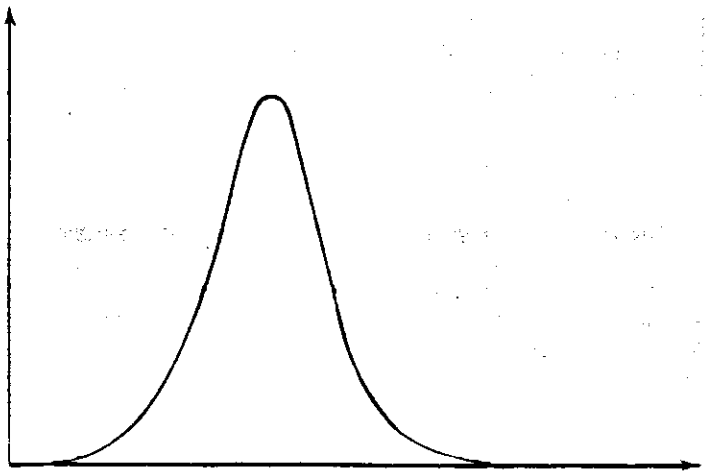


FIGURE 1

Cette courbe est parfaitement symétrique de part et d'autre de la note la plus fréquente ***(le mode).*** *Elle représente une distribution dite "normale" car elle caractérise à peu près la forme de données naturelles comme la mesure de la taille ou du poids, de la fréquence cardiaque... L'apparition d'une distribution normale, d'un point de vue probabiliste, s'explique par l'existence de facteurs de variations nombreux, dont les influences sont faibles et de même importance, et qui sont indépendants l'un de l'autre. La distribution normale, en pédagogie comme en psychologie, est souvent conçue comme idéale. Si les résultats à une épreuve de sciences physiques ou à un test d'intelligence se distribuent normalement, certains*

auront tendance à interpréter ce constat comme une justification naturelle des différences de niveaux scolaire et/ou intellectuel. On voit bien le danger d'un amalgame entre le comptage statistique et l'interprétation idéologique. En effet, rien ne prouve que les facteurs qui interviennent dans la variation des résultats à une interrogation scolaire ou à un test soient indépendants les uns des autres, équivalents en importance et chacun de faible influence! D'autre part, il faut savoir que si les résultats à un test se distribuent "normalement" ce peut être par la volonté de son auteur qui a construit le test pour obtenir cette forme, ou qui a opérer des transformations mathématiques des notes (voir infra). Une distribution normale peut recouvrir plutôt une manipulation statistique qu'un phénomène naturel !

Il arrive que la courbe soit dissymétrique, le mode se situant à gauche ou à droite (fig. 2).

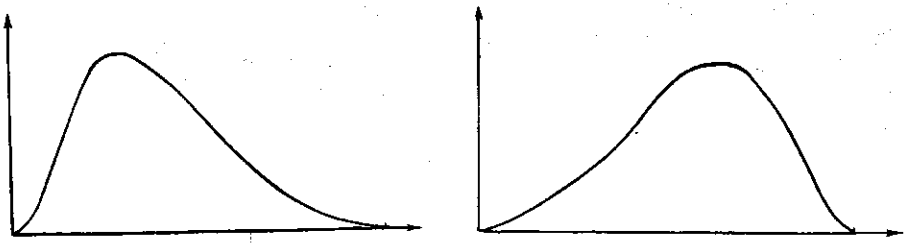


FIGURE 2

Les courbes en I et en J sont franchement dissymétriques (fig. 3). La première traduit un échec massif, lié à un niveau de difficulté trop élevé par rapport au public ou à une sévérité de correction extrême. Le J, au contraire, traduit la réussite et pour certains l'idéal pédagogique à atteindre, à la condition qu'il ne traduise pas le laxisme du correcteur !

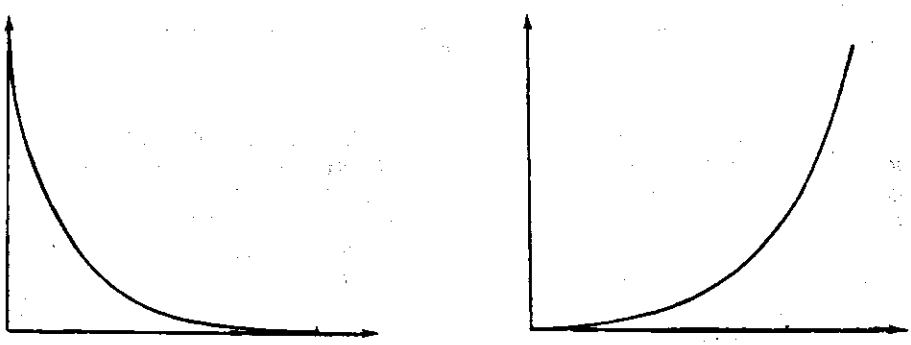


FIGURE 3

Les distributions à deux ou plusieurs modes (fig. 4) indiquent une hétérogénéité de la population. Par exemple on a reporté ensemble les résultats à une épreuve d'athlètisme d'un groupe d'hommes et d'un groupe de femmes.

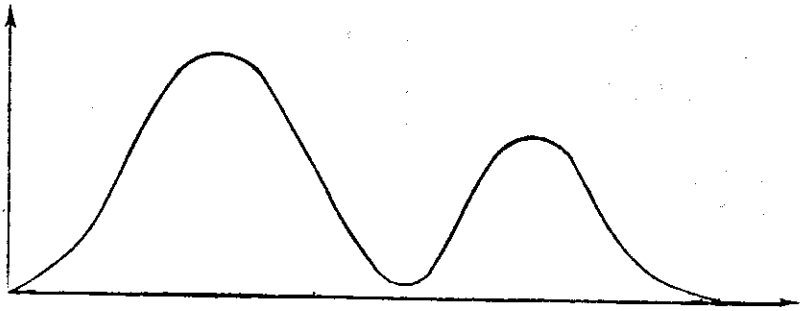


FIGURE 4

La distribution en U (fig. 5) s'interprète comme la juxtaposition d'une courbe en I et d'une courbe en J quand on met ensemble un groupe très performant et un groupe très faible.

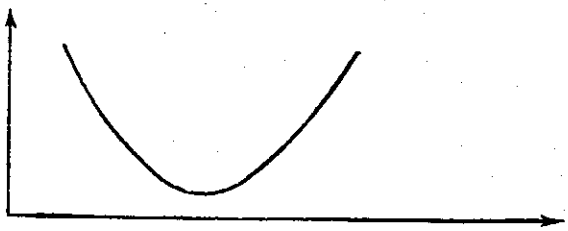


FIGURE 5

Revenons à note liste de notes brutes trop inorganisée pour l'instant. La première transformation utile pour structurer l'information consiste à élaborer un **tableau d'effectifs**. Observons le tableau ci – dessous (tab. 1).

TABLEAU 1

N	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311
E	1	0	0	0	0	1	1	2	0	0	0	2	0	2	2	3	0	3	1	1	2	2	1	0	1	3	0	0	1	0	0	1
EC	1	1	1	1	1	2	3	5	5	5	5	7	7	9	11	14	14	17	18	19	21	23	24	24	25	28	28	28	29	29	29	30

N : notes brutes obtenues au test
E : effectifs ou nombre de fois qu'une même note brute apparaît
EC : effectifs cumulés ou somme propressive des effectifs.

Dans ce tableau, on a ordonné les notes de la plus petite à la plus grande et on a compté le nombre d'apparitions de chacune des notes. On voit que les notes brutes vont de 280 à 311 signes copiés dans les 4 minutes de travail, soit une étendue de (311 – 280) = 1 32 signes. On remarque que le nombre de fois où une note a été attribuée varie de 0 à 3. La valeur 3, nombre d'apparitions le plus élévé, est présente 3 fois. Il y a donc 3 modes dans cette distribution, soient les notes 295, 297, 305.

Bien évidemment, le fait qu'une note n'ait pas été attribuée ne veut pas dire qu'elle ne le sera jamais! A une autre occasion, ou avec d'autres sujets, on obtiendrait une dispersion à chaque fois un peu différente.

Pour rendre l'information plus visuelle, il est d'usage de construire un ***histogramme en bâtons ou en rectangles****. L'histogramme en rectangle est permis lorsque les notes sont continues, c'est-à-dire lorsqu'on peut théoriquement admettre que toutes les notes intermédiaires pourraient être attribuées. Par exemple, on admet qu'entre 1m50 et 1m51 on peut donner les valeurs 1m506, 1m5067...*

Dans notre test-exemple, on considère que l'unité de mesure élémentaire est un signe, lettre ou chiffre, complet: mesurer des moitiés de signes, des dixièmes ou des millièmes de signes n'aurait aucun sens. On se place donc dans un cas où les notes

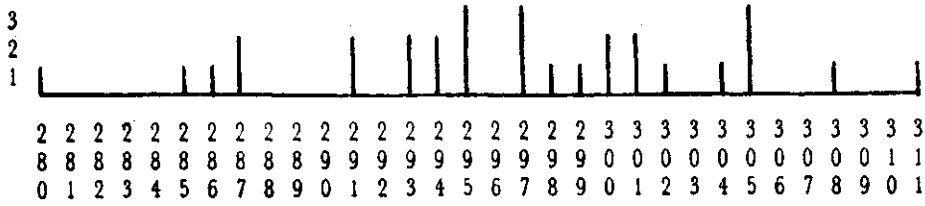


FIGURE 6

sont discontinues, séparées les unes des autres par un espace, ce qui est une indication pour la réalisation d'un histogramme en bâtons. Dans ce type de représentation, l'effectif de chaque note se traduit par un "bâton" dont la hauteur lui est proportionnelle (fig. 6).

Une forme commence à se dessiner, mais elle conserve quelques imperfections. Il y a des vides dans la distribution des notes, il y a toujours les 3 modes. Dans ces conditions, il est difficile de décrire clairement la tendance générale des résultats.

Afin de faire ressortir une forme interprétable d'une liste ou d'un histogramme de notes brutes qui "ne parle pas" nettement, on utilise la technique du regroupement en classes. Il y a deux moyens de constituer des classes. Le premier consiste à découper la distribution des notes en parts égales, une classe étant formée tous les 2, 3, 4... n points. Le second alimente les classes en effectifs de notes selon des proportions déterminées a priori.

L'étendue en points d'une classe, ou son effectif sont fixés de façon arbitraire. Pratiquement, divers critères peuvent intervenir selon les besoins, ou les habitudes, ou les préférences... Si on souhaite faire ressortir une classe centrale, on doit choisir une étendue de la distribution ou des pourcentages d'effectifs qui donneront un nombre impair de classes. Par exemple, sur un ensemble de 5 classes, la troisième sera la classe centrale...

Le regroupement en classes introduit une "condensation" des observations. En effet, on diminue le nombre d'éventualités. Par exemple, si je transforme un continuum de 60 notes brutes en 5 classes, j'aurai réduit le nombre d'éventualités de 60 à 5 soit de 12 fois. La condensation en classes entraîne donc une perte d'informations dans la mesure où des notes différentes mais situées dans la même classe sont considérées comme identiques. On peut aussi critiquer le fait que deux notes qui se suivent pourront être situées dans des classes différentes alors que deux notes éloignées marquant les deux bornes d'une même classe ne seront pas différenciées. Quand on travaille avec des classes, il est donc imprudent de vouloir comparer des écarts relatifs entre individus: 2 individus appartenant à une même classe peuvent être plus distants que 2 individus appartenant à deux classes différentes.

Un autre critère de choix dans l'étendue ou l'effectif des classes est relatif à l'objectif de faire disparaître les vides et les irrégularités que l'on obtient toujours avec de petits échantillons de sujets. Cette démarche se justifie en considérant que ces irrégularités sont liées aux facteurs de variations "locaux", spécifiques aux caractéristiques des sujets de l'échantillon. C'est une considération acceptable puisqu'avec un autre échantillon de sujets, les irrégularités et les vides apparaîtront à d'autres endroits. Le but du regroupement en classes, c'est de révéler la forme des résultats, la règle ou la tendance générale qui se dilue, qui se cache dans le hasard de la constitution des échantillons. Le nombre idéal de classes devient alors celui qui préserve un nombre d'éventualités suffisant c'est-à-dire discriminatif des performances tout en faisant ressortir la forme générale de la distribution des niveaux de performance dans la population considérée.

Utilisons la première manière – découpage en parts égales de l'échelle des notes brutes – avec les données dont nous disposons. On peut décider de regrouper les notes de 5 en 5 points. Nous obtiendrons ainsi 7 classes. Chaque classe se définit par ses limites – note inférieure et note supérieure –, par sa valeur centrale – note du milieu –, et par son effectif. On peut transformer l'effectif en fréquences (tab. 2).

Tableau de regroupement en classes :

TABLEAU 2

Classes	1	2	3	4	5	6	7
Notes – Valeurs limites	280 – 284	285 – 289	290 – 294	295 – 299	300 – 304	305 – 309	310 – 314
Notes – Valeur centrale	282	287	292	297	302	307	312
Effectifs	1	4	6	8	6	4	1
Fréquences	0,03	0,13	0,20	0,26	0,20	0,13	0,03

A partir du tableau d'effectifs, on peut réaliser un histogramme en bâtons (fig. 7).

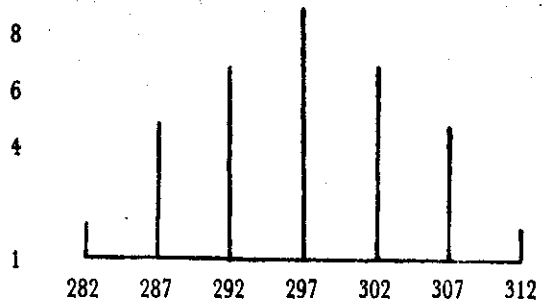


FIGURE 7

Cet histogramme est différent du premier obtenu avec les données brutes. On voit se dessiner une tendance centrale forte, nettement plus avantageuse que les 3 modes de l'histogramme précédent, puisque 8 sujets sur les 30 écrivent 297+ ou –2 signes en 4 minutes. La forme de la distribution obtenue est probablement proche de celle qui existe dans la population parente des candidates aux postes de travail envisagé: elle en représente une estimation. Pour exprimer cette forme de manière indépendante de

l'échantillon, on considère les fréquences. Par exemple, 6 sujets de l'échantillon ont copié 292 signes + ou −2. Ces 6 sujets représentent une certaine proportion de l'échantillon égale à 6 : 30 soit 0,20 (ou 20%). La fréquence calculée sur l'échantillon est une estimation de la fréquence de cette vitesse d'écriture qui devrait exister dans la population parente. On peut donc représenter la forme de la distribution de la rapidité d'écriture dans l'ensemble des candidates potentielles en utilisant seulement les fréquences présentées sous la forme d'un **polygone** (fig. 8).

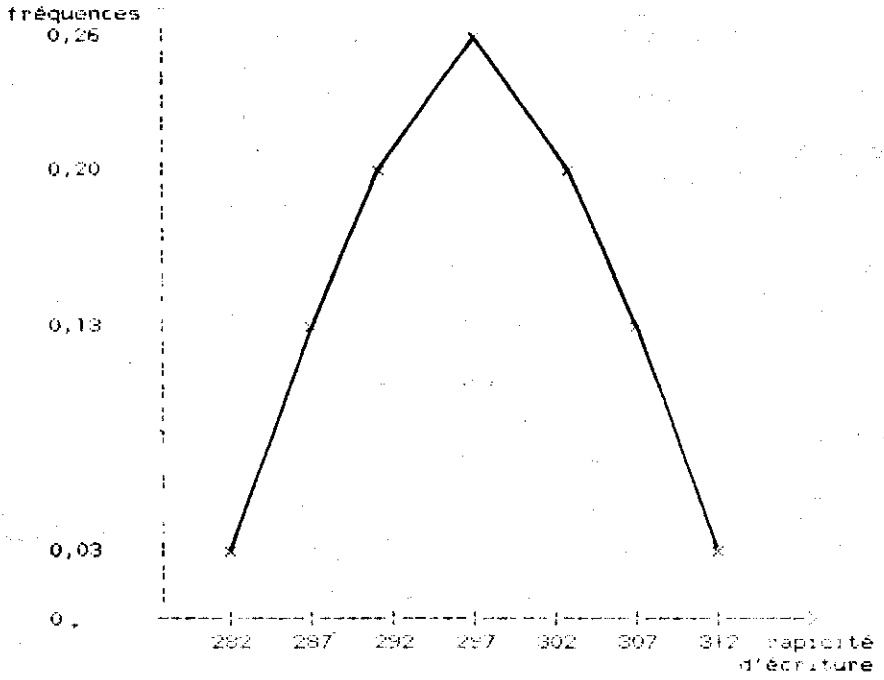


FIGURE 8

Si l'échantillon est bien représentatif, on peut énoncer que 3% des candidates potentielles ont une vitesse de 282 + ou −2 signes pour 4 minutes de travail, que 13% ont une vitesse de 287 + ou −2 signes...

Si l'on considère que cette mise en forme des résultats convient, on peut alors passer à la cinquième et dernière étape, la construction de l'étalonnage. Sinon, on peut regrouper en classes en jouant sur les effectifs.

Il existe différents types de regroupements par les effectifs. On peut utiliser des formes en référence à la courbe de GAUSS. Dans ce cas, on dit qu'il y a ***normalisation de la distribution*** *en 5, en 7, en 9, en n classes. Mais en pédagogie, nous avons déjà remarqué que rien ne permettait de conclure à une distribution statistiquement normale des connaissances. On peut postuler par exemple qu'il serait*

souhaitable que les résultats d'un enseignement se rapprochent d'une courbe en J avec un maximum d'élèves qui aient assimilé correctement les contenus et un minimum d'élèves éprouvant des difficultés d'apprentissage. Tout dépend également de l'objectif poursuivi. Prenons l'exemple des pratiques de sélection. Soit on décide de ne retenir que les meilleurs - sélection de cadres supérieurs par exemple... -. Dans ce cas, il faut arriver à bien différencier les performances dans les scores élevés. On choisira alors une courbe dissymétrique, bien étalée sur sa partie droite. Soit on décide d'éliminer les plus mauvais. La forme des résultats devra alors différencier les performances faibles et sera dissymétrique, plus étalée dans sa partie gauche cette fois. Il existe également des ***distributions "rectangulaires"*** *dans lesquelles on égalise les effectifs de chaque classe. Ainsi, un* ***quartilage*** *est formé de 4 classes, chacune regroupant 25% de la population... Les distributions rectangulaires sont critiquées parce qu'elles différencient de manière illusoire les sujets moyens. Mais nous avons vu que le choix d'une distribution relève de l'arbitraire et que selon les circonstances, on peut être amené à préférer telle ou telle forme (tab. 3). Il est vrai que la recherche d'une distribution normale par le psychologue est souvent très active, parfois à la limite de la "salade statistique" car ce type de distribution ouvre la possibilité de tous les calculs que permet une courbe mathématiquement définie.*

Tableau des pourcentages par classe en fonction du mode de découpage choisi :

TABLEAU 3

distributions rectangulaires

- quartilage : 4 classes à 25%
- décilage : 10 classes à 10%
- centilage : 100 classes à 1%

distributions normales

- en 5 classes : 6,7%; 24,2%; 38,2%; 24,2%; 6,7%
- en 7 classes : 4,8%; 11,1%; 21,2%; 25,8%; 21,2%; 11,1%; 4,8%
- en 9 classes : 4%; 6,6%; 12,1%; 17,6%; 19,6%; 17,6%; 12,1%; 6,6%; 4%
- en 10 classes : 4,3%; 5,6%; 9,6%; 13,9%; 16,6%; 16,6%; 13,9%; 9,6%; 5,6%; 4,3%
- en 11 classes : 3,6%; 4,5%; 7,7%; 11,6%; 14,6%; 16%; 14,6%; 11,6%; 7,7%; 4,5%; 3,6%

Pratiquement, pour construire la distribution à partir des effectifs, on s'appuie sur les effectifs cumulés. Reprenons le tableau d'effectifs tiré de notre exemple (tab. 1). Nous allons l'utiliser pour construire une distribution rectangulaire puis une distribution normalisée.

N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3
	8	8	8	8	8	8	8	8	8	8	9	9	9	9	9	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	1	1
	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1
E	1	0	0	0	0	1	1	2	0	0	0	2	0	2	2	3	0	3	1	1	2	2	1	0	1	3	0	0	1	0	0	1
E	1	1	1	1	1	2	3	5	5	5	5	7	7	9	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	3
C															1	4	4	7	8	9	1	3	4	4	5	8	8	8	9	9	9	0

Construction d'un décilage. Dans un décilage, il y a 10 classes, chacune comportant 10% de l'effectif total. Puisque nous avons 30 sujets, dans chaque classe il devra y avoir 3 sujets. Les effectifs cumulés du décilage seront donc: 3, 6, 9, 12, 15, 18, 21, 24, 27 et 30. On met en correspondance les effectifs cumulés du décilage et les effectifs cumulés des notes brutes pour déterminer les limites des classes. On obtient (tab. 4) :

TABLEAU 4

Classes	1	2	3	4	5	6	7	8	9	10
Notes brutes	280 286	287 290	291 293	294	295 296	297 298	299 300	301 302	303 304	305 312

Construction d'une distribution en 5 classes. Les effectifs à attribuer à chacune des 5 classes pour un groupe de 30 sujets sont les suivants: 2,01; 7,26; 11,46; 7,26; 2,01. Une fois cumulés ils deviennent: 2,01; 9,27; 20,73; 27,99; 30. On met en correspondance les effectifs cumulés de la distribution normalisée et les effectifs cumulés des notes brutes pour déterminer les limites des classes. On obtient (tab. 5) :

TABLEAU 5

Classes	1	2	3	4	5
Notes brutes	280 - 285	286 - 293	294 - 299	300 - 307	308 - 312

On peut ainsi construire autant de formes de distributions que l'on veut et le lecteur peut s'entraîner à l'élaboration d'un quartilage, ou à une normalisation en 7 classes à partir de nos données ou des siennes propres.

5 - La dernière étape de l'élaboration d'un test est la construction de l'étalonnage. Le travail a été fait puisque l'étalonnage c'est tout simplement la distribution reconstituée des notes brutes selon une forme choisie à priori.

Nous disposons donc de 2 étalonnages: un décilage et une échelle normalisée en 5 classes.

Comment utiliser un étalonnage? Imaginons que les jours suivants une autre occasion nécessite de mesurer la rapidité d'écriture pour 3 personnes, A, B et C. On leur fait passer le test et on le corrige en suivant rigoureusement les consignes. Les notes brutes sont les suivantes: A a 283 points, B a 295 points, C a 307 points.

Grâce aux étalonnages on va pouvoir situer les trois performances par rapport au groupe de référence et donc par rapport à la population parente (tab. 6).

TABLEAU 6

	Décilage	N.S.	Echelle normalisée	N.S.
Sujet A	1ère classe	1	1ère classe	1
Sujet B	5ème classe	5	3ème classe	3
Sujet C	10ème classe	10	5ème classe	5

N.S. : note standardisée

Comment lire un tel résultat? Prenons le sujet B. Il se situe dans la 5ème classe du décilage, ce qui veut dire que sa performance est meilleure que 40% des performances et moins bonne que 50% des performances. Il se situe également dans la 3ème classe de l'échelle normalisée, ce qui veut dire que sa performance est meilleure que 30,9% des performances et moins bonne que 30,9% des performances. On peut dire que la note de B se situe comme une note médiane ou moyenne par rapport aux autres notes. On situe ainsi la performance individuelle sur une échelle de mesure à n degrés.

Il faut faire attention toutefois à une lecture trop rapide de certains résultats donnés sous la forme de notes standardisées. Le numéro d'ordre des classes est souvent pris comme une note d'un second degré, une note standardisée résultante d'un étalonnage. Dire que quelqu'un a eu 7 au test ne veut pas dire 7/10 ou 7/20! Il ne faut pas assimiler notes standardisées et notes scolaires. La note standard doit être accompagnée du type d'étalonnage pour pouvoir s'interpréter. Dans notre exemple, A obtient 5 sur le décilage et 3 sur l'échelle normalisée pour une même performance. Il arrive également que certains auteurs attribuent la classe 1 à la performance la plus élevée et la classe n à la performance la plus basse. Afin d'éviter des lectures erronées, il y a intérêt à bien comprendre au préalable de quelle manière l'étalonnage est présenté.

Pour revenir à notre recrutement imaginaire et le conclure, il reste à raconter le principe de décision. Nous disposons des résultats de 30 candidates. Il faut déterminer le seuil de points qui indique une vitesse adaptée au type de travail. Pour établir ce seuil, le test sera proposé à un groupe d'opératrices jugées performantes par leur encadrement. Elles obtiennent toutes les notes standardisées 4 et 5 sur l'échelle en 5 classes, ou 8, 9 et 10 sur le décilage. C'est donc 1/3 des candidates qui seront évaluées comme aptes à effectuer ce travail du point de vue de la rapidité et une liste de 10 candidates pourra être proposée à l'entreprise cliente.

Suite à l'élaboration de notre test-exemple, on comprend mieux ce qu'est un test. Toutefois, il ne faut pas se laisser prendre au piège de la technicité. On peut – et on doit – rechercher la forme de la distribution et le type d'étalonnage qui correspondent le mieux aux besoins de l'évaluation pressentie, au type de décision à prendre, aux caractéristiques de la population... Mais l'usage d'une statistique sophistiquée est un leurre si le test ne mesure pas ce pourquoi il est construit et si l'échantillon, le groupe de référence n'est pas représentatif de la population visée. Concrètement cela veut dire que les premières étapes, élaboration du contenu et de la forme, constitution de l'échantillon, sont primordiales. Un test ne s'improvise pas. C'est une affaire de spécialiste qui nécessite apprentissage et expérience. Pour l'utilisateur, cela pose le problème du choix du test parmi les catalogues des maisons d'édition de ce type de matériel. C'est là l'objet du chapitre suivant.

CHAPITRE 11

COMMENT IMPLANTER UN TEST PEDAGOMETRIQUE ?

Lorsqu'on n'est pas informé d'un domaine de connaissances et qu'on souhaite l'aborder, une stratégie première consiste à développer des conduites par essais et erreurs. Cette stratégie s'impose pour des domaines entièrement nouveaux, mais ces tâtonnements expérimentaux sont coûteux en temps, en énergie et peuvent mener au découragement et à l'abandon si les résultats tardent à se concrètiser. Pour éviter cette forme de gaspillage aux formateurs et aux enseignants désireux d'utiliser des tests, nous allons essayer de faire le tour des conditions de mise en oeuvre de l'épreuve la mieux adaptée aux besoins.

Dans la représentation commune, l'usage des tests est uniquement rattaché à la pratique professionnelle des psychologues. C'est là une conception de l'usage des tests beaucoup trop restrictive qu'il serait bon de dépasser afin que les enseignants et les formateurs, les responsables des services formation des entreprises, les conseillers en formation continue et les techniciens des centres de bilan personnel et professionnel... puissent disposer d'outils d'évaluation scientifiquement élaborés, pratiques et fiables sous réserve du respect de leurs règles d'application: les tests pédagométriques.

Sous le terme générique de tests pédagométriques, nous regroupons l'ensemble des tests qui peuvent être utilisés dans le champ de la formation initiale et continue. Ce sont des épreuves étalonnées qui introduisent la mesure et une métrique dans le domaine de la pédagogie.

Dans ce vaste ensemble, on peut dessiner trois catégories de tests en fonction de leur objectif principal.

En premier lieu, et parce qu'ils sont les plus nombreux, viennent les **tests de connaissances**. Le but d'un test de connaissances est de situer le niveau d'une personne par rapport à une population parente, par rapport à un groupe d'appartenance. Leur utilisation se situe à la fin d'un cycle de formation ou d'une unité d'enseignement. Par exemple, avant de recruter des hôtesses de l'air, on doit vérifier leur niveau de connaissance de la langue anglaise. A la fin de l'année scolaire, pour faire le bilan de son enseignement, le professeur de physique de la classe de première S veut situer le niveau de connaissances de ses élèves par rapport à l'échelle des niveaux de l'ensemble des premières S. Le professeur d'électricité du Lycée d'enseignement professionnel, pour préparer son programme de seconde année, veut situer le niveau des connaissances théoriques et le niveau des maîtrises pratiques de ses élèves à la fin de leur première année par rapport à l'échelle des niveaux des élèves de L.E.P. de première année d'électricité. Pour attribuer une qualification à un nouvel embauché, le chef du personnel d'une entreprise de mécanique générale veut situer précisément le niveau de connaissances professionnelles du nouvel agent par rapport aux échelles de niveaux de chaque catégorie de personnel... Les tests de connaissances se subdivisent selon le domaine de connaissance évalué. Il peut s'agir des disciplines de base et on aura les tests de niveau de français, de mathématiques, de physique, d'anglais, d'espagnol... Mais aussi il existe des tests plus spécialisés dans des domaines professionnels ou techniques et on trouve les tests de dessin technique, de connaissances juridiques, de connaissances mécaniques, de connaissances médicales... Une variante, inspirée directement de l'évaluation formative, développe des "tests centrés sur les objectifs" ou encore des "tests de maîtrise". Ces épreuves visent...
"à révéler le répertoire comportemental d'un sujet ou d'un groupe eu égard aux objectifs de l'apprentissage... L'instrument de cette évaluation doit donc fournir des indications qui sont directement interprétables en fonction d'objectifs" (TOURNEUR). Dans cette alternative, la compétence n'est pas appréciée par le classement de la performance par rapport à un groupe de référence, mais comme l'importance du chemin parcouru vers l'atteinte d'un objectif, ou comme distance qui reste à parcourir vers cet objectif. Ainsi, par exemple, l'enseignant ou le formateur peuvent connaître suite à une phase pédagogique la proportion d'élèves ou de stagiaires qui possèdent le niveau minimum de connaissances pour suivre correctement la phase suivante. Dans ce genre d'épreuves le choix de la forme de la distribution des résultats n'a aucune importance puisqu'il est tout à fait envisageable, et même souhaitable pour l'enseignant dans la mesure où cela fait partie de son métier, que l'ensemble des stagiaires ou des élèves atteignent tout ou partie des objectifs à 100%. Très certainement utiles pour réguler le cours d'un enseignement ou d'une formation, ces épreuves n'ont pas de valeur générale pour tous les élèves puisque les objectifs poursuivis présentent un grand degré de variabilité d'un enseignant à un autre. Au mieux seront-elles des épreuves communes à l'échelon d'un établissement lorsqu'un travail en équipes pédagogiques s'y développe effectivement. Toujours dans le cadre d'une évaluation formative, des établissements ont fait avancer la pratique des **profils pédagogiques**. Un profil pédagogique est un document qui trace un portrait analytique et global d'un élève ou d'un stagiaire à partir d'observations recueillies sur une certaine durée et portant sur un ensemble diversifié de dimensions. Ainsi, seront pris en compte les disciplines scolaires, des savoirs tranversaux, des compétences pratiques, des qualités personnelles, des capacités sociales, des

intérêts et des activités de loisir... (voir un exemple planche n° 1). Bien que différent des formes de l'évaluation scolaire et professionnelle classique, le profil ne remplace pas le test standardisé dans la mesure où aucune comparaison à un groupe de référence n'est proposée. Toutefois, rien n'empêche de construire un profil psychologique ou pédagogique à partir d'une série de résultats obtenus à la passation d'une batterie de tests (voir exemples planches n° 2, 3). Les profils présentent l'avantage de regrouper l'essentiel de l'information évaluative. Ils permettent de situer l'ensemble des niveaux de capacités (comparaison aux normes du groupe de référence) d'une part et d'apprécier l'homogénéité des performances (comparaison des niveaux entre eux) pour faire ressortir points forts et points faibles relatifs d'autre part.

En second lieu, on peut regrouper les **épreuves pronostiques**. Ce sont des tests qui donnent des prédictions. Par exemple, ils indiquent la prévision d'un durée d'un apprentissage, la prévision du niveau de connaissance qui sera atteint... Dans la mesure où ces prévisions s'accompagnent d'une probabilité, leurs résultats peuvent se formaliser en termes de **risque pédagogique**. Ces tests s'utilisent en amont d'une formation. Ils peuvent servir en direction des élèves et des stagiaires pour individualiser les parcours pédagogiques, et en direction des enseignants et des formateurs pour moduler le rythme d'une progression, pour différencier des objectifs, pour établir les contenus. Ils peuvent servir en direction des établissements pour aider à mettre sur pied l'organisation pédagogique générale, pour constituer des groupes de niveaux, des groupes en soutien...

En troisième lieu, on distingue les **épreuves diagnostiques**. Ces tests ont pour objectif de participer à la compréhension des difficultés éprouvées au cours d'un apprentissage. Dans l'idéal, les épreuves diagnostiques doivent présenter deux qualités :
"... fournir des informations sur les lacunes individuelles et communes des élèves d'un niveau défini... [et]... permettre une ***analyse étiologique des erreurs"*** (KETELE et PAQUAY). Ce type d'analyse nécessite que les processus de la démarche de l'élève puissent être mis en lumière. Ces épreuves s'utilisent en cours de formation, si possible dès d'un problème d'apprentissage est perçu, tant au niveau du groupe classe qu'au niveau individuel. Les résultats obtenus, dans le cas de difficultés générales au groupe, permettent d'envisager de quelle manière il convient de réajuster le niveau d'exigence magistrale, d'abaisser le niveau de complexité et/ou le niveau d'abstraction, de ralentir le rythme de l'introduction des notions nouvelles, de multiplier les entraînements fonctionnels nécessaires à l'enracinement des nouveaux acquis, de modifier le mode de transmission du savoir ou du savoir-faire... Dans le cas de difficultés vues sous l'angle individuel, les résultats du test fournissent des indications précieuses pour définir et pour amorcer un programme de rééduction et d'aide psychopédagogique.

A l'intérieur de ces trois grandes catégories de tests pédagométriques, il est possible de considérer deux facteurs de différenciation: le mode de passation et le mode de réponse. La passation est soit collective, soit individuelle; le mode de réponse est soit le questionnaire à choix multiple (le Q.C.M.), soit l'expression libre de la réponse ouverte. Les tests de niveau sont plutôt des Q.C.M. collectifs. Les épreuves pronostiques se présentent sous des formes variées et les épreuves

diagnostiques sont nécessairement individuelles et à expression libre puisqu'il s'agit de repérer de quelle manière le sujet fonctionne mentalement pour accomplir la tâche.

Le principal point de discussion concerne l'utilisation des Q.C.M. (voir un exemple de Q.C.M. planche n° 4). Ce mode de réponse est très pratique puisqu'il permet une correction rapide par simple application d'une grille et même l'utilisation efficace de l'ordinateur qui donne tous les pourcentages de réussite souhaités: par question, par groupes de questions, par élève, par sous-groupes d'élèves (filles et garçons, redoublants et non-redoublants...), par classe et groupes de classes... Toutefois, plusieurs reproches ont été formulés.

En premier lieu, lorsqu'on demande de choisir une réponse parmi trois ou quatre propositions, la réussite liée à un choix au hasard conserve une probabilité consistante. Pour diminuer ce risque, différentes techniques sont utilisées. Par exemple, on augmente le nombre de propositions, ou bien la consigne demande au sujet de parier un nombre de points sur sa réponse, ou bien la bonne réponse ne figure pas toujours dans les propositions et une case vide est mise à disposition pour que le sujet puisse y inscrire cette réponse dans les cas où cela est nécessaire... Dans d'autres cas, et par construction, les propositions de réponses sont élaborées systématiquement selon certains critères. Par exemple, chacun des choix comprendra une réponse juste, une réponse approximative, une réponse présentant une information inexacte, une réponse correspondant à une lourde ignorance, une réponse "je ne sais pas" (BUJAS). Ou bien, les propositions de réponses erronées reprennent des erreurs réellement commises par les sujets parents au cours d'épreuves préalables où les questions du test sont posées de manière ouverte (PASQUIER, T.P.6.). Une autre critique des Q.C.M. consiste à dire qu'un mode de réponse simple (cocher une case) ne peut rendre compte que de modes de pensée également simples. En fait, la manière de présenter la question peut demander la mise en route de raisonnements élaborés. Ainsi, on peut proposer deux faits et le sujet doit décider de l'exactitude de ces faits et du type de relation qui existe entre eux. D'autre part il est évident qu'un mode de réponse direct n'a pas de relation avec la longueur du détour opératoire et/ou cognitif à mettre en oeuvre pour élaborer la solution: si on demande de calculer 34 234 679 x 879 763 903 il est clair que le temps de calcul n'a pas de rapport direct avec le mode de réponse qui consiste à cocher une case! Si on pose le problème des lapins de FIGONACCI :
"un couple de lapins, nés à la même date 0, donne naissance, à partir du deuxième mois de son existence, à un couple chaque mois. Ces nouveaux couples suivent la même loi de reproduction; calculer mois après mois l'évolution du nombre de couples...", là encore, le mode de réponse n'a pas de rapport avec la qualité du raisonnement! Toutefois, le Q.C.M. ne renseigne pas sur la démarche suivie par le sujet. Il est commode et rapide pour situer un niveau, mais il est limité pour apprécier les qualités expressives et personnelles des individus. Conduire un diagnostic avec un Q.C.M. s'arrêtera au mieux à l'élaboration d'un profil, et l'analyse des erreurs ne sera pas possible. Un autre danger concerne l'emploi éventuellement abusif des Q.C.M. en tant que base de décision pour l'orientation scolaire et professionnelle qui conduirait probablement à un apprentissage des tests à des formes de bachotage et même de "trafic de documents" (GUILBERT) ou de "cours spécialisés" autour des tests. L'évaluation formelle prend alors le pas sur l'éducation ce qui conduit à un

appauvrissement des capacités rédactionnelles et d'une manière plus générale à un appauvrissement culturel des élèves, des stagiaires, des candidats à des postes de travail...

Donc l'usage des tests pédagométriques n'est pas réservé aux seuls psychologues. Ce n'est pour autant que l'on puisse se permettre de faire n'importe quoi, n'importe comment. Pour un enseignant qui n'a pas l'habitude de manipuler l'outil évaluatif qu'est un test, nous conseillons de commencer par les test de connaissances. Dans un second temps, il pourra se familiariser avec les épreuves pronostiques à finalité préventive et ensuite avec les épreuves diagnostiques à finalité curative, les plus délicates à utiliser. Bien entendu, l'enseignant ou le formateur ne doivent pas se priver des aides qui leur sembleraient nécessaires à certains moments. Ils peuvent déjà travailler en équipe avec leurs collègues qui souhaitent également utiliser les tests, l'idéal étant bien entendu que la pratique de l'évaluation à l'aide des tests pédagométriques s'inscrive dans le cadre d'un projet d'établissement. Ils peuvent aussi poser leurs questions au psychologue scolaire, à l'orthophoniste, au rééducateur, au conseiller d'orientation, au psychologue du travail lorsque ces personnels sont présents dans leurs institutions et qu'ils ne se montrent pas hostiles à l'usage des épreuves pédagométriques. Enfin, le plus simple est souvent d'adresser un courrier à la maison d'édition.

Rappelons les principes de base de l'utilisation des tests, principes qui doivent être suivis tant par les débutants que par les utilisateurs confirmés. En effet, au fur et à mesure des applications, les utilisateurs ont tendance à personnaliser la manière de procéder. Des parties de consigne jugées inutiles sont escamotées ou résumées, les temps indiqués sont rallongés pour que tous les sujets puissent terminer, l'examinateur ne peut s'empêcher de fournir des aides, il va se servir de ses propres critères de correction en place de ceux de l'auteur du test, il délaisse les étalonnages, se contentant d'avoir une idée du niveau atteint... tant et si bien que l'aspect métrique du test qui fait son intérêt et sa spécificité se dilue dans la fantaisie. Il est donc nécessaire de ressourcer régulièrement sa pratique aux principes de base.

Le premier de ces principes est le **respect du copyright**, c'est-à-dire de la protection légale de l'oeuvre. Concrètement, cela veut dire qu'il est rigoureusement interdit de photocopier les épreuves. Cette protection légale protège l'éditeur et aussi les droits des auteurs qui ne sont rémunérés que sur les ventes réalisées. Le matériel de test comprend généralement un manuel, des livrets d'exercices et parfois des feuilles ou des profils de résultats ainsi que des grilles de correction.

Le second principe est celui de la **confidentialité**. Cela ne veut pas dire que si l'on fait passer des tests pédagométriques on devient membre d'une secte ou d'une société secrète! La confidentialité joue sur deux niveaux. D'abord, le contenu du test ne doit pas être divulgué à quiconque. Elaborer un test demande deux à trois années de travail, parfois plus. Divulguer son contenu à tous vents revient à le rendre inutilisable car tous les sujets susceptibles de passer l'épreuve l'apprendront par coeur avant la passation. Dans le même ordre d'idée, les sujets qui ont passé le test ne seront jamais informés des réponses correctes, ils ne recevront pas leur "copie" corrigée en retour. D'une part ils peuvent avoir à repasser le même

test à une autre occasion, d'autre part ils pourraient rendre publiques les réponses correctes. Il faut aussi prendre le soin de fournir des feuilles de papier brouillon si nécessaire, facilement reconnaissables (de couleur par exemple) et toutes les ramasser à la fin de la passation. Ces précautions peuvent sembler ridicules ou dérisoires, mais elles déterminent la durée opérationnelle du test. La confidentialité joue à un second niveau en ce sens que **les résultats au test appartiennent au sujet qui l'a passé.** En conséquence, il n'y aura pas d'annonce publique des résultats individuels. Cela n'empêche pas **l'obligation de restitution.** Concrètement, les résultats seront restitués à chacun des sujets qui ont passé le test. Dans le cas de sujets mineurs et seulement dans ce cas, les parents devront être invités à venir prendre connaissance des résultats en même temps que leur enfant. D'une manière générale, **la passation du test ne doit pas porter préjudice au sujet**, ce qui au fond est la règle de conduite la plus importante à respecter.

La seconde série de principes concerne les aspects pratiques de la mise en oeuvre des tests pédagométriques. Une fois qu'il se sera procuré le matériel, le futur examinateur s'empressera de... ne rien entreprendre! Il convient d'abord de procéder à la **lecture approfondie du manuel.** Au cours de cette lecture, il devra repérer les différentes parties du texte: présentation théorique de l'outil évaluatif, la consigne, la méthode de correction et le barème de notation, le ou les étalonnages, la ou les études réalisées avec ce test...

Il est tout à fait utile dans un premier temps de **faire le test sur soi-même.** C'est la meilleure manière d'acquérir une vue de l'intérieur du test, de connaître à fond son contenu (et à l'occasion de repérer ses propres lacunes), et donc de savoir exactement ce qu'il mesure (part du vocabulaire actif et du vocabulaire passif, part de l'usage et des accords... dans une épreuve de français par exemple).

Dernière précaution avant que de lancer une classe entière sur le test: **réaliser quelques essais avec des petits groupes.** Ceci est très important pour plusieurs raisons. On limite ainsi les conséquences des erreurs éventuellement commises par l'apprenti testeur. Et les erreurs possibles sont nombreuses! Citons les plus classiques. Le contrôle du temps de passation des épreuves données en temps limité n'est pas évident: on oublie de noter l'heure de départ (ou de fin) ou de déclencher (ou d'arrêter) le chronomètre. On ne contrôle pas que chaque sujet a bien travaillé toutes les pages du livret d'exercices. On n'a pas prévu un nombre suffisant de livrets. On donne toutes les pages d'un coup alors qu'il faut les donner une par une. On interprète mal la consigne. On laisse commencer les sujets impulsifs avant le signal de départ. On se laisse avoir au sentiment par des sujets qui vous extorquent des éléments de réponse. On s'aperçoit a posteriori que des sujets ont pu communiquer entre eux. On applique mal le transparent de la grille de correction sur la page d'exercice. On commet des erreurs de calcul en faisant l'addition des bonnes réponses. On se trompe dans la lecture de l'étalonnage... Les essais préalables permettent également à l'apprenti-testeur de se familiariser avec l'ensemble des phases de la passation, de s'entraîner à dire correctement la consigne quand elle est orale, de repérer les cas limites à la correction, de comprendre le système de passage de la note brute à la note standard...

Le jour de la "première" est arrivé. Il est maintenant essentiel de ne plus se tromper si l'on veut obtenir des résultats valables. **La préparation matérielle se fait avant la passation.** C'est là un principe simple mais important si l'on ne veut pas être empoisonné toute une passation par un détail matériel. Pouvoir répondre à tous les imprévus de la passation suppose que l'on se soit libéré d'un maximum de contraintes avant cette passation. Rappelons que la préparation matérielle participe au confort général du sujet et à la validité des résultats: il est essentiel que les sujets ne puissent communiquer entre eux. Si nécessaire il vaut mieux scinder un groupe trop important par rapport à la salle et aux tables disponibles en deux sous-groupes et réaliser deux passations.

On ne commence pas un test brutalement. **Il faut accueillir et mettre à l'aise les sujets.** Au cours de la phase d'accueil, si l'on observe que l'un des sujets ne se trouve pas dans son assiette, qu'il paraît comme malade ou trop fatigué, qu'il donne les signes d'une perturbation importante, il est préférable de différer la passation pour cette personne. Pour les autres, il convient de faire baisser le niveau d'anxiété qui accompagne toute situation stressante. Il ne s'agit pas que tout le monde se calme au point de s'assoupir, mais il faut essayer de détendre l'atmosphère, de dialoguer avec les individus, de s'intéresser à eux, de leur montrer qu'on les prend en considération en tant que personne... D'un autre côté, il faut arriver à ce que chacun puisse donner le meilleur de lui-même et pour cela il convient de susciter l'intérêt, de maintenir un état de vigilance, de concentration et d'attention suffisant. En bref, il convient de parvenir à donner une tonalité positive au stress de toute façon nécessaire pour mobiliser les ressources physiologiques et intellectuelles qui permettront de surmonter l'épreuve.

Il faut **informer les sujets du pourquoi de la passation du test pédagométrique.** Cette information se fait à deux niveaux. Sur le plan institutionnel, il convient de dire les raisons qui fondent la décision d'utiliser un test, d'exposer le processus décisionnel dans lequel s'inclut la passation du test, de situer les enjeux. Bien entendu, ce discours sera adapté aux caractéristiques du groupe de sujets. On ne s'adresse pas de la même manière à des enfants de grande section d'école maternelle et à des cadres supérieurs d'une entreprise, même si le fond du message est identique: le test est utilisé pour faire le point sur un certain type de connaissances, de savoir-faire... Ce sont les objectifs de l'institution qui vont différer à chaque fois: bilans d'orientation, de classement, de recrutement, de qualification professionnelle, de formation initiale ou continue... La connaissance de l'enjeu social a un effet direct sur la motivation et la qualité de la mobilisation des investissements qui en découlent et il est donc important que tous les sujets aient un même niveau d'information sur cet aspect de l'évaluation. Le deuxième niveau concerne le test pédagométrique lui-même. Il y a lieu d'expliquer, sauf indication contraire de l'auteur, ce que mesure le test, quels savoirs ou savoir-faire il interroge, quelles capacités il stimule... Cette connaissance préalable du domaine des compétences qui vont être explorées mobilise chez les sujets les aires de la mémoire, les représentations, les expériences acquises et les capacités fonctionnelles en rapport avec ce domaine. Cela permet à chacun d'atteindre plus rapidement son meilleur rendement. Ce temps d'information préalable restera d'une courte durée, sa fonction principale étant de planter le décor, de faire comprendre la signification et les finalités de l'évaluation.

Une fois que le groupe se retrouve correctement installé, à la fois détendu et concentré, et suffisamment informé on va donner la consigne du test. **La consigne du test pédagométrique doit être suivie à la lettre.** Ce principe est l'un des plus délicats à appliquer. La consigne est là pour faire comprendre au sujet ce qu'on attend de lui. Si le sujet ne comprend pas cette consigne, il ne pourra pas faire le test et on ne pourra pas conclure quant au domaine évalué. Par exemple, à un Q.C.M. de connaissances en dessin industriel, un sujet met systématiqueent des croix dans toutes les cases. Il est évident que ce sujet n'a rien compris au fonctionnement du mode de réponse mais on ne peut rien dire quant à ses connaissances en dessin industriel. Pour éviter ce genre de mésaventure, en règle générale, il est conseillé de contrôler la compréhension de la consigne à partir des premières réponses produites par les sujets. Si l'incompréhension est générale, il est préférable d'arrêter tout le monde et de redonner la consigne en s'appuyant peut-être plus largement sur l'exemple ou les exemples qui le plus souvent accompagnent l'énoncé. Sinon, on réexplique individuellement, toujours sur l'exemple donné, jamais sur une question du test. Notons bien, et c'est cela qui est le plus difficile à respecter, que les explications données au sujet, que la consigne soit orale ou qu'elle soit écrite, sont relatives à *la manière de répondre et seulement à la manière de répondre.* Concrètement, l'examinateur ne doit jamais donner une explication, ni même une indication ou un indice, qui puissent présenter un caractère facilitateur, mettre sur la voie pour trouver le résultat correct, ou induire l'application de tel ou tel raisonnement... En régle générale, l'examinateur adopte **une attitude de neutralité bienveillante.** En clair, il fait tout pour placer les sujets examinés dans les meilleures conditions matérielles possibles et les amener aux dispositions psychologiques les plus favorables à la résolution d'une tâche, -dimension de la bienveillance-, mais il ne fait rien qui puisse guider vers une réponse correcte, -dimension de la neutralité-. Cette attitude est souvent difficilement acceptable par les enseignants et les formateurs qui débutent dans l'usage des tests dans la mesure où ils ne supportent pas de ne pouvoir intervenir quand ils voient certains de leurs élèves ou de leurs stagiaires éprouver des difficultés ou se tromper massivement. C'est pourquoi nous conseillons, à chaque fois que cela est possible, qu'il y ait permutation des maîtres pour les moments d'évaluation sommative et pour faire passer les tests de niveaux plus particulièrement.

La règle générale de neutralité bienveillante souffre deux exceptions au moins. Dans le premier cas le problème ne relève pas principalement du classement de la performance d'un sujet par rapport à celles des sujets d'un groupe de référence représentatif d'une population parente. Cette circonstance concerne les tests à finalités diagnostiques. Certes, en première approche ces épreuves permettent de situer une performance parmi d'autres, mais elles doivent permettre d'aller plus loin et d'explorer les facteurs qui ont amener à ce niveau. Pour ce faire, il est essentiel que la procédure de passation prenne une tournure interactive. L'objectif de l'examinateur devient la mise en évidence des *processus de pensée* qui conduisent à produire telle ou telle réponse, à réussir dans un type de problème mais à échouer dans un autre... Reprenons un exemple donné par KOCHER. On fait lire le texte suivant repris de CLAPAREDE à un enfant de 9 ans :

"Le vigneron achève la taille de la vigne pour que le raisin vienne en abondance. Il a travaillé hier soir jusqu'à une heure tardive. Ce matin la première lueur du

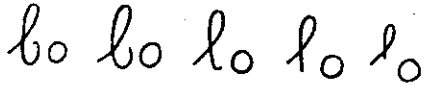
jour le retrouve déjà à l'ouvrage. Brave vigneron, ton labeur mérite d'être récompensé".

L'enfant lit de la manière suivante :

"Le violon achè... acheté la tèle... Le violon acheté. La tèle de la vigu... vigue pour que les raisons vinne en ado... adon... abondance. Il a tar... tava... il a travaillé hier. Jusqu'au... jusqu'à une heure tard... très tardi... jusqu'à une heure tardive. Ce matin le peur... premier... le premier feu... f... l... lu... lu... leur du jour le ret... le retou... le retrou... ve déjà à l'ou... à l'orage dave... darave vi... vije... ron ton lad... eur mirite d'être r... réc... récompense... récompensé".

Le classement de la performance lexique sur la base d'un nombre de mots correctement lus par minute n'aurait ici aucune signification. Le problème à résoudre est l'analyse qualitative des erreurs et la recherche de leurs sources afin de trouver le remède le mieux approprié en termes de rééducation de la lecture. Le centrage se déplace du résultat vers la démarche intellectuelle de son élaboration. Par rapport à cet objectif, l'examinateur ne peut rester neutre et il va poser les questions en changeant leurs formes et leurs contenus jusqu'à ce qu'il obtienne l'expression du processus. Restons dans le domaine de la lecture pour regarder un bel exemple de recherche de processus que nous livre LEFAVRAIS. Il travaille avec un enfant de 7;3 ans et s'aperçoit, parmi d'autres erreurs que la syllabe "bo" est lue "cho". Il imagine, à titre d'hypothèse, qu'un fragment privilégié retient l'attention spontanée du sujet. Pour trouver ce fragment, LEFAVRAIS réduit progressivement la boucle de la lettre "b".

L'enfant lit "cho" à chaque nouvelle graphie. La lettre "b" est alors diminuée dans sa partie inférieure.



Une fois la boucle disparue, l'enfant ne lit plus rien. Enfin, on propose à l'enfant des parcelles quelconques de la syllabe et...
"comme je m'y attendais, seuls, ceux contenant la partie croche provoquent une réponse qui est uniformémemt "cho"...".

D'autres expériences vont montrer que cet écolier, quelle que soit la graphie qui lui est soumise, lettre, syllabe ou mot, va y rechercher un fragment qui pour lui seul veut "dire" quelque chose. Il s'est inventé un code phonologique personnel basé

sur une appréhension parcellaire du signifiant, source d'erreur première qui va, si elle n'est rééduquée à temps, se cristalliser et empêcher tout apprentissage correct et efficace. On voit bien à travers cet exemple que l'analyse étiologique des erreurs devient possible en rapprochant la nature de l'erreur et le mode de fonctionnement individuel. A ce moment là, le besoin de mesure cède le pas devant le besoin d'analyse. Il ne s'agit plus de projeter un jugement de valeur mais de comprendre ce qui freine ou empêche l'efficacité individuelle afin de rechercher un mode d'intervention pédagogique mieux approprié à chaque cas et qui soit susceptible de lever les inhibitions cognitives et/ou affectives, de remédier aux dysfonctionnements, aux incompréhensions, aux mauvaises habitudes, de combler les lacunes relatives aux règles, aux contenus, à l'application généralisée des règles sur n'importe quel contenu...

La deuxième exception à la règle de neutralité bienveillante concerne la passation des épreuves qui visent la mesure d'un potentiel d'apprentissage. Ces tests font mettre en oeuvre une phase où un apprentissage est proposé aux sujets. Au cours de cette phase, on donne des indications non sur la manière de répondre mais sur la manière de trouver la réponse, sur la manière de raisonner... Puisque le but devient alors la réalisation d'un apprentissage, on vise une acquisition du sujet, voire un début de modification de son fonctionnement, de ses conduites, de ses comportements, de ses stratégies, de ses modes de représentations du réel et de traitement des informations...

Même si l'aide est standardisée dans sa procédure, l'examinateur ne peut rester neutre et les aspects relationnels, affectifs et motivationnels vont prendre le dessus sur les aspects formels de l'évaluation le temps de cet apprentissage.

Une fois la passation terminée, si l'on veut utiliser l'étalonnage, il faut **suivre scrupuleusement les critères de correction et les barèmes de notation.** Il arrive aux enseignants apprentis-testeurs d'éprouver de grandes résistances pour appliquer un mode de notation qu'ils n'ont pas élaborer eux-mêmes. Il faut comprendre que la standardisation de la correction est une nécessité absolue pour la validité de la mesure: le mètre ne peut pas être élastique et on ne peut se permettre de changer d'unité ou d'échelle de mesure au gré des fantaisies individuelles. Tous les codages universels ont un caractère conventionnel et donc arbitraire que nous sommes obligés d'accepter: si je veux rouler en voiture automobile, je dois accepter le code de la route en son état; si je veux exercer le métier de géomètre ou de mettreur, je dois accepter le système métrique tel qu'il est... Alors, il faut être très attentif au début de la pratique pédagométrique, se forcer à la rigueur quand cela s'avère indispensable. C'est la dernière condition d'une utilisation correcte des étalonnages. Une fois le résultat brut connu, il peut être rapporté à l'étalonnage et transformé en note standardisée.

La dernière étape concerne l'interprétation du résultat. Il faut vraiment se méfier d'une tendance qui nous habite tous, celle qui consiste à considérer les choses d'une manière absolue. **Il convient de relativiser le résultat.** Tout d'abord, **si la réussite traduit bien une compétence liée à un ensemble de capacités, l'échec ne traduit pas une incompétence.** Ce n'est pas parce que j'ai échoué dans une tâche aujourd'hui que j'échouerai forcément demain! L'histoire des sports, des sciences et des techniques, des arts, des hommes politiques... multiplient les exemples de ce

qu'on appelle des tentatives. Le résultat obtenu doit être réinséré dans les différents contextes qui l'ont généré. Il faut, par exemple, regarder si le résultat est concordant avec les performances habituelles de chacun des sujets ou bien s'il présente le caractère d'une exception; il faut voir les conditions paticulières au jour de la passation; il faut vérifier que les contenus scolaires ont été bel et bien enseignés en cours d'année... Il y a toujours intérêt à discuter du résultat avec chacun des sujets qui fourniront bien souvent une bonne part des éléments de relativisation.

A travers tout cela, on voit que la passation des tests pédagométriques est possible par tout enseignant et formateur sans forcément suivre un cursus de formation spécialisée à la condition toutefois de respecter un certain nombre d'indications générales. Il faut principalement se rappeler que si l'application d'un test demande beaucoup de rigueur, cette rigueur doit se développer avec finesse et intelligence. D'autre part, l'usage du test doit correspondre à un besoin et l'épreuve sera choisie en fonction de ce besoin. Mais comment choisir?

CHAPITRE 12

COMMENT CHOISIR UN TEST PEDAGOMETRIQUE ?

L'emploi d'une épreuve standardisée répond à un besoin d'évaluation métrique. Il y a souci de mesurer la performance d'un élève, d'un stagiaire ou d'un groupe-classe en situant cette performance sur l'échelle des résultats qu'obtiendrait une population parente appréhendée au travers d'un échantillon qui en donne une image représentative. Ce souci de la mesure répond à un besoin d'objectivité qui se satisfait dans le recours à une norme externe au pédagogue, norme référée à une population beaucoup plus large que celle de la classe ou du groupe de stagiaires en formation. Il répond également à un besoin de relativisation des performances individuelles par leur comparaison à un ensemble de mesures beaucoup plus étendu. Il répond encore, d'une manière plus générale, à un besoin d'évaluation sur deux plans: évaluer le niveau de chaque individu; évaluer les effets de la pédagogie mise en oeuvre car il ne faut jamais oublier qu'un niveau scolaire ou professionnel est toujours le produit d'un type d'enseignement scolaire ou professionnel. L'usage d'une métrique aide le formateur et l'enseignant à se distancier de l'imprécision des impressionnismes à la source des allégations gratuites car non vérifiées du genre "le niveau baisse actuellement"... Toutefois, l'ensemble de ces besoins ne suffit pas pour choisir tel ou tel test pédagométrique. En fonction des paramètres qui définissent chaque situation locale, il convient de prendre en compte successivement chacun des différents critères exposés ci-après. L'ordre de présentation de ces critères a été organisé, mais il n'a rien d'absolu et on pourrait envisager des alternatives à cette présentation. D'autre part l'ordre retenu correspond à un degré de généralité élevé qui ne peut prétendre s'harmoniser avec toute situation locale, aux caractéristiques particulières. Il est certain que parfois la faiblesse des budgets attribués fera que le critère "prix du matériel" prendra de l'importance. Si on se place dans le cadre d'une recherche, le critère "études réalisées avec le test" passera au premier plan...

Sur un plan pratique, les descriptions des épreuves et les renseignements relatifs à leurs utilisations figurent dans les catalogues fournis par les maisons spécialisées dans l'édition des tests. Les grandes entreprises et les principaux organismes de formation nationaux ou privés élaborent souvent leurs propres tests, mais ils ne les diffusent commercialement. On ne peut donc pas parler de ces travaux. Il existe, en France, deux maisons d'édition principales :

- les Editions du Centre de Psychologie Appliquée
  48, avenue Victor Hugo
  75783 PARIS CEDEX 16

- les Etablissements d'Applications Psychotechniques
  6 bis, rue André Chénier
  92130 ISSY-LES-MOULINEAUX

Dans chaque catalogue, les tests sont classés, par section pour les Editions du Centre de Psychologie Appliquée (le C.P.A.), et par rubrique pour les Etablissements d'Applications Psychotechniques (les E.A.P.). Dans le catalogue du C.P.A., on trouve des tests pédagométriques dans les sections II "aptitudes", III "connaissances", IV "épreuves cliniques". Dans celui des E.A.P., il faut rechercher dans les rubriques 3 "batteries factorielles aptitudes diverses", 6 "tests de connaissances", 7 "tests d'aptitudes et de connaissances scolaires". (voir en annexe la liste de ces tests).

Les pages de ces catalogues ne se présentent pas de la même manière (voir planches n° 5 et n° 6). Considérons une page du catalogue du C.P.A. Dans la colonne de droite figurent, dans le cadre du haut, la section et le titre abrégé du test, dans le cadre médian le nom en clair du test et dans le cadre du bas l'année de l'édition. En pleine page, on voit figurer de bas en haut le nom en clair du test, le nom de l'auteur ou des auteurs, le ou les éditeurs, suivi des rubriques suivantes: matériel, administration, notice technique et bibliographie. Une page du catalogue des E.A.P. comprend en cadre haut le titre en clair du test et le nom de l'auteur, son sigle et la rubrique. Ensuite, on a les divisions suivantes: description, utilisation, statistiques, présentation et bibliographie.

Afin d'aider à la détermination du choix de l'épreuve, nous proposons donc d'appliquer la grille de définition de critères suivante. On peut distinguer un premier groupe de critères vus comme **les déterminants principaux du choix** car en rapport direct avec l'objet de l'évaluation.

Le premier déterminant concerne le *domaine de connaissances* à évaluer. Veut-on évaluer les connaissances requises pour apprendre à lire? les connaissances en français, en anglais, en espagnol, en mathématiques, en hydraulique, en médecine, en droit...? En règle générale, l'indication du domaine de connaissances est comprise dans le titre de l'épreuve. En cas de doute, on peut essayer de lever l'ambiguïté en recherchant des précisions dans le texte de présentation. On peut ainsi sélectionner une première liste d'épreuves.

Le second déterminant est relatif *au groupe de référence et à la population parente.* Il s'agit de voir à quel type de population s'adresse le test. Une épreuve

d'anglais étalonnée avec des élèves de collèges ne conviendra pas pour des lycéens. Une épreuve prédictive d'un apprentissage de la lecture étalonnée avec des enfants de grande section de maternelle ne convient pas pour des adultes immigrés... Notons ici avec insistance que l'utilisation d'un test avec un individu ou un groupe de sujets pour lequel il n'a pas été étalonné n'a aucun sens: on ne peut comparer que ce qui est comparable. Mieux vaut ne pas utiliser de test quand on ne trouve pas un étalonnage correspondant au besoin. Ces indications figurent au C.P.A. dans la partie "administration" et aux E.A.P. dans les parties "utilisation" et "statistiques". Dans les deux cas, on regardera les deux indications suivantes: "Age d'application" et "Etalonnage(s)". Parfois, ces indications sont insuffisantes et nécessitent de rechercher des indications complémentaires dans le texte de présentation de l'épreuve. Par exemple, l'indication "Etalonnage: Population de 300 sujets pour la Forme Adulte" ne dit pas si ces sujets sont des hommes ou des femmes, s'ils appartiennent à telle catégorie sociale ou à toutes les catégories, s'ils appartiennent à telle tranche d'âges ou à toutes les tranches d'âges... Autre exemple, pour un test de niveau scolaire on lit: "Age d'application: classes primaires" et "Etalonnage: enfants français". Il faut rechercher dans la notice technique pour savoir que les étalonnages concernent "le niveau des élèves au début de l'année scolaire". Ce test n'est donc pas utilisable pour dresser un bilan de fin d'année scolaire et si tel est mon objectif, je dois choisir une autre épreuve. Suite à l'application de ce critère on peut réduire la liste des épreuves virtuellement intéressantes en éliminant celles qui ne proposent pas un étalonnage approprié.

Le troisième et dernier déterminant principal du choix concerne *l'objectif de l'utilisation*. Nous rejoignons ici la catégorisation des épreuves utilisée au chapitre précédent pour différencier les tests pédagométriques en fonction de leurs indications: tests de niveau, tests pronostiques, tests diagnostiques. Il faut donc trouver une épreuve qui corresponde exactement à notre objectif. A priori, une épreuve de niveau ne peut pas servir à la formulation d'un pronostic et encore moins à l'établissement d'un diagnostic. Toutefois, certains tests sont mixtes et, par exemple, en fonction d'un niveau atteint à l'issue d'un cours donnent une probabilité de réussite dans le cours suivant. Ou bien, une épreuve de niveau de lecture est prévue pour faciliter la catégorisation des erreurs... Seul le C.P.A. indique une utilisation globale à la rubrique "Administration, Indications". Il faut donc lire avec le plus grand soin le texte de présentation pour en retirer toutes les utilisations possibles selon les auteurs. La rubrique "bibliographie" peut se révéler utile quand on a la chance de pouvoir consulter les articles ou les ouvrages donnés en référence. A travers les études présentées dans les textes, on peut ainsi prendre connaissance de la manière dont le test a été mis en oeuvre dans telle ou telle circonstance et surtout par rapport à tel ou tel objectif. Suite à l'application de ce dernier critère principal, on peut une fois encore réduire le nombre des épreuves possibles.

Après cette sélection, trois cas de figures sont envisageables. Premier cas, aucune épreuve ne correspond à mon projet sur tout ou partie des critères appliqués: domaine de connaissance, groupe de référence et objectif de l'utilisation. Il n'y a pas de solution immédiate. A plus long terme deux alternatives consistent soit à formuler son besoin auprès des maisons d'édition pour qu'elles étudient la question, soit à mettre en chantier une épreuve adaptée à ses besoins en jouant sur ses propres

ressources. Dans cette optique, il y aura nécessité de suivre une formation spécialisée aux statistiques appliquées à l'évaluation. Second cas, une seule épreuve ressort. Comme on n'a pas le choix, il faut l'essayer! Troisième cas, plusieurs tests sont en compétition. On a le luxe de l'embarras du choix. Il est alors possible de faire jouer les autres critères. Pour plus de facilité, si la question du budget ne se dresse pas comme un obstacle insurmontable, nous conseillons au futur utilisateur de se procurer les "specimen-set" qui en général comprennent un test, une feuille de réponse si elle existe, un manuel d'application avec les étalonnages, les grilles ou les corrigés n'étant pas fournis. On peut ainsi voir et toucher le test, faire quelques bouts d'essais... Et surtout, on peut lire le détail du manuel d'application pour y rechercher les renseignements qui ne figurent pas forcément sur les pages du catalogue.

Le second groupe de critères à mettre en oeuvre dans le choix d'un test pédagométrique concerne les **qualités métrologiques de l'épreuve standardisée**. Le test pédagométrique est un instrument de mesure et il doit présenter toutes les qualités des autres instruments de mesure.

La première de ces qualités est la **validité**. Le *vocabulaire de la psychologie* de PIERON donne la définition suivante de la validité d'un test :
"La validité d'un test exprime le degré de liaison entre le rendement du sujet dans le test et son rendement dans une autre activité, que le test est censé prévoir".

Dit autrement, un test doit réellement mesurer ce que son auteur dit qu'il mesure. Depuis PIERON le concept de validité s'est différencié et on peut maintenant considérer au moins trois types de validité.

La validité de contenu (on dit aussi la validité logique, on la validité de construction, ou encore la validité interne) consiste à examiner le contenu du test pour vérifier qu'il est conforme à ce que l'épreuve doit mesurer. Si je contrôle une épreuve de calcul pour des élèves de cours moyen, je dois conclure sur trois plans: (1) toutes les questions portent bien sur les contenus mathématiques qui devraient être enseignés au cours moyen d'après les instructions officielles qui déterminent ces contenus; (2) l'ensemble des questions proposées est représentatif de toutes les questions possibles que l'on pourrait poser pour évaluer le niveau de connaissances en mathématiques d'un élève du cours moyen; (3) la manière de poser les exercices et le mode de réponse retenus ne risquent pas d'interférer avec l'évaluation des connaissances en mathématiques proprement dites. La première vérification (1) est relativement simple à opérer par comparaison entre contenu du test et contenu des programmes officiels. La seconde (2) se révèle plus délicate puisqu'elle pose le problème de la rationnalisation d'un choix. En principe, l'auteur du test devrait expliciter les limites de l'univers du contenu appréhendé et surtout les critères qu'il a retenus pour opérer ses choix et/ou ses exclusions. Lorsque cela n'est pas précisé, c'est à l'utilisateur à abstraire les critères de choix et à les apprécier à leur juste valeur. Enfin, la troisième vérification (3) est encore plus problématique et suite à un examen attentif du test peut demander de procéder à des vérifications expérimentales sur le terrain. Si le test se présente comme un Q.C.M. avec peu d'éventualités de réponses, ou s'il contient des propositions qu'il faut juger vraies ou fausses, il faut vérifier que les élèves qui n'ont pas peur de répondre au hasard

ne seront pas avantagés par rapport aux élèves prudents qui, lorsqu'ils doutent ou ne savent pas, préfèrent s'abstenir. Si tel est le cas, il est à craindre que le test mesure plus une attitude devant l'incertitude qu'un niveau de mathématiques proprement dit. Dans le même ordre d'idée, mais dans le domaine de l'orthographe, rien ne garantit a priori qu'un sujet qui réussit bien dans la détection et la correction d'erreurs dans un texte qui lui est proposé soit bon en dictée ou rédige sans faire de fautes. La nature de la tâche est donc également à prendre en considération afin de vérifier si les situations proposées par le test sont représentatives des situations dans lesquelles le contenu ou le comportement évalués peuvent apparaître.

La validité concourante (on dit encore validité empirique, ou validité externe) correpond à la forme de validité évoquée par PIERON: on cherche à établir une comparaison entre les résultats que donnent le test et un critère extérieur. Si je reprends l'exemple du test de mathématiques pour le cours moyen, je vais comparer dans une dizaine de classes la correspondance statistique entre les résultats obtenus au test et la moyenne des notes ou appréciations attribuées en mathématiques depuis le début de l'année. Si cette correspondance est correcte, je peux en tirer deux conclusions: le test donne bien un niveau de connaissances en mathématiques; il permet d'établir ce niveau en quelques heures pour toute une classe au lieu d'attendre une année. La liaison statistique entre le test et le critère est la plupart du temps exprimée par un **coefficient de corrélation**. Sans entrer dans le détail technique, pour comprendre la lecture de ces coefficients il suffit de savoir que plus leur valeur est proche de +1 et plus la liaison test-critère est étroite, plus elle est proche de 0 et plus la liaison est faible et à la limite inexistante, plus elle est proche de –1 et plus la liaison est étroite mais de sens inverse: une forte note au test accompagne une faible note au critère et réciproquement. Pour valider un test de cette manière, on peut aussi prendre pour critère un autre test dont la validité a été nettement démontrée.

La validité prédictive correspond à la valeur pronostique du test. Le test est valide dans la mesure où il permet de faire un pronostic c'est-à-dire de formuler une prévision relative à une réussite future dans une quelconque activité. La marge d'erreur qui accompagne le pronostic du test doit être nettement inférieure à celle qui accompagne une prévision intuitive, "au pifomètre". Dans le cadre de la pédagométrie les tests pronostiques sont intéressants en amont des formations, en début de chaque cursus, afin d'évaluer les chances de réussite qui déterminent le risque pédagogique de chaque élève ou stagiaire. Le pédagogue peut ainsi se centrer d'une manière très attentive sur les élèves à risque afin de réguler au mieux leurs apprentissages d'une manière préventive. Lorsque le risque est important on peut d'emblée compléter l'évaluation par des épreuves permettant de poser un premier diagnostic des difficultés d'apprentissage.

Une fois que l'on a pu vérifier la validité du test, on peut passer à l'examen de la seconde qualité métrologique d'un test: **la fidélité**. Toute mesure physique et à plus forte raison toute mesure pédagogique est entâchée d'erreurs. Un test fidèle est un test fiable, c'est-à-dire un test dont la marge d'erreur de mesure est la plus faible possible. Concrètement, deux, trois, n... mesures consécutives du même objet doivent idéalement donner des résultats sensiblement identiques. L'erreur de mesure

peut provenir de l'examinateur qui procède à la mesure et/ou de l'instrument qu'il utilise et/ou d'une transformation de l'objet mesuré. La variation de mesure entre examinateurs, dans le cadre des épreuves standardisées, est théoriquement nulle lorsque deux conditions sont remplies: tout examinateur suit et applique rigoureusement les indications portées dans le manuel de l'épreuve; ces indications ont bien prévu tous les cas de figure possibles. Pour apprécier l'erreur de mesure liée à l'instrument-test, trois techniques sont utilisées.

Le *test-retest* consiste à donner l'épreuve deux fois à un même groupe de sujets, à un intervalle de temps donné. On compare, toujours à l'aide de coefficients de corrélation, les classements des sujets à ces deux passations. Si le degré de correspondance entre les deux classements est élevé, supérieur à .80 par exemple, on dira que le test est fidèle dans la mesure où il situe les sujets les uns par rapport aux autres de la même manière: il y a constance de la mesure. On n'obtient jamais de fidélité parfaite pour plusieurs raisons. Entre les deux passations, les circonstances externes ont pu évoluer: les sujets seront plus fatigués, moins motivés par le test qu'ils voient pour la deuxième fois, ils auront peut-être subi le contre-coup de diverses perturbations ou d'expériences récentes... Si les deux passations sont rapprochées dans le temps, on peut craindre des effets d'apprentissage du test différents d'un sujet à l'autre...

La seconde technique consiste à utiliser ce qu'on appelle des *formes parallèles.* Deux formes parallèles d'un même test sont deux formes identiques: même nombre de questions; même présentation et même mode de réponse. Seuls les contenus diffèrent légèrement, mais ils restent de même nature et d'un même niveau de difficulté. Il est possible d'appliquer successivement ces deux formes à un même groupe de sujets puis de comparer les deux classements obtenus. Si leur corrélation est bonne, on peut conclure à l'équivalence de ces deux formes. L'usage de ce dispositif diminue sensiblement les éventuels effets d'apprentissage du dispositif test-retest et permet de réduire le temps qui sépare les deux passations. Toutefois, on peut s'attendre à ce que l'équivalence parfaite, et en conséquence la fidélité parfaite, entre des formes parallèles reste une utopie dans la mesure où il n'est pas facile d'harmoniser rigoureusement et sur tous les plans les questions deux à deux. La troisième et dernière technique est dite des *questions paires-questions impaires* (on peut voir aussi "item pairs-item impairs" ou "méthode pair-impair" ou "fidèlité split half"). Lorsque le test comporte un nombre suffisant de questions, et après l'avoir fait passer à un groupe de sujets, on va le diviser en deux sous-tests composés le premier de toutes les questions numérotées avec des nombres pairs, le second de toutes les questions numérotées avec des nombres impairs. On obtient ainsi deux séries de notes et on peut évaluer la corrélation des classements des sujets obtenus à partir de chacun des deux sous-tests. Si la correspondance entre ces classements est bonne, on peut conclure à une bonne homogénéité de l'épreuve ce qui est aussi un indice de sa fidèlité. On aura donc intérêt à choisir parmi les tests possibles, ceux dont la fidèlité est la plus forte.

La dernière qualité métrologique concerne la **sensibilité du test** (on dit aussi la finesse discriminative). Pour PIERON,
"Un test est dit avoir une plus ou moins grande finesse discriminative selon qu'il peut classer les individus en un nombre plus ou moins grand d'échelons".

Un test qui classerait tous les individus au même niveau, c'est-à-dire que tous les sujets obtenant la même note, ne serait d'aucune utilité! Toutefois, une finesse trop importante, qui différencierait chacun des membres de la population parente, nécessiterait autant d'échelons que de personnes! Ce qui peut s'appliquer à la mesure de la performance sportive comme le passage au centième voire au millième de seconde représente des échelles de distinction illusoires dans le domaine de l'évaluation pédagogique. La question du nombre d'échelons rejoint tout le problème du choix de la forme de la distribution et de son découpage en classes évoqué au chapitre dixième. Le choix du test par rapport à cette question de la sensibilité est tout à fait relatif. Autant il est important que l'épreuve soit dans l'absolu valide et fidèle, autant le choix de la finesse discriminative répond au projet et aux besoins personnels de l'examinateur.

Après avoir appliqué les deux premiers groupes de critères, il reste peut-être encore un choix possible entre deux ou plusieurs épreuves. Nous pourrons alors faire jouer quelques **critères pratiques**.

Le **mode de passation** est à considérer par rapport à la gestion du temps dont on dispose, au nombre de sujets à examiner, à l'objectif du test. Peu de temps disponible, beaucoup de sujets, une évaluation de niveau sont des indications pour des tests collectifs. Un temps par sujet plus étendu, une évaluation diagnostique sont des indications pour des tests individuels.

Les possibilités d'**analyse qualitative** sont bien sûr à considérer quand l'examinateur a le souci de mieux comprendre les processus de réponses des sujets. Dans ce cas les épreuves mixtes niveau-diagnostic sont intéressantes.

Le **mode de réponse** intervient également dans cette optique d'allier le quantitatif et le qualitatif pour explorer les relations entre démarche, type d'erreurs et résultat. Il est sûr qu'un Q.C.M. ordinaire n'apporte que peu d'éléments à analyser qualitativement. En passation collective, les questions ouvertes ou le sujet doit élaborer et écrire sa réponse seront plus riches de ce point de vue malgré qu'elles prennent un peu plus de temps à corriger. De toute manière, l'analyse qualitative vaut la peine d'être conduite quand derrière le test pédagométrique des mesures pourront se décider en ce qui concerne l'amélioration de l'enseignement, son individualisation, la mise en route de rattrapages, de soutiens, voire de rééducations.

On peut également prendre en compte le **type de capacité intellectuelle** qui entre en jeu par rapport à la nature de la tâche. Il faut voir alors si le test fait simplement appel à une restitution de connaissances qui mobilise uniquement la mémoire, ou bien s'il demande des activités de résolution de problèmes impliquant des raisonnements plus ou moins complexes et sophistiqués, ou encore des activités d'expression et de création. Le choix dépend bien entendu du projet global de l'évaluateur et ce sur quoi il veut focaliser son attention.

Les **références bibliographiques** peuvent aider à déterminer son choix en fonction d'exemples d'études et de recherches réalisées avec tel ou tel test pédagométrique.

Enfin, et il ne faut pas l'oublier, le critère économique, c'est-à-dire le rapport qualité-prix, sera à prendre en considération. A qualité de service rendu égale, il est justifié de choisir le test le moins onéreux. Dans les autres cas, je crois que l'essentiel est que le test puisse correctement répondre aux besoins exprimés par l'évaluateur. Acheter une Ferrari pour aller acheter son pain au coin de la rue constitue un gaspillage de ressources et d'énergie tout autant que de creuser les fondations de sa maison à la petite cuiller! Lorsqu'on compare les prix, il faut différencier trois postes différents: coût du manuel par rapport au nombre de pages; coût du matériel réutilisable qui ne s'achète qu'une fois; coût du matériel consomptible, à renouveler pour chaque utilisation nouvelle. Si l'on pratique de nombreuses passations, ce dernier poste devient le plus important à considérer.

Une fois que tous les critères ont été appliqués, s'il reste encore des épreuves en concurrence, alors il suffit de tirer au sort l'heureuse élue. Conrètement, il arrive plutôt assez souvent qu'aucune épreuve ne résiste à une sélection rigoureuse. Dans ce cas, on peut se permettre de rechercher le test le plus proche du besoin d'évaluation pédagométrique, en gardant à l'esprit toutefois qu'on ne peut guère transiger avec les critères principaux. On peut à la rigueur se contenter d'un instrument de mesure aux qualités métrologiques un peu douteuses, mais il n'est pas possible de jouer avec la signification essentielle de l'évaluation d'un domaine de connaissances défini, par rapport à un groupe de référence correspondant du groupe de sujets, dans un objectif affiché à l'avance. S'il subsiste quelque doute qui freine l'engagement dans un choix, il ne faut pas oublier que le meilleur recours reste le questionnement de l'éditeur et de l'auteur du test pédagométrique.

CONCLUSION

Du bon usage des tests en pédagogie :

Tout du long de la partie du livre consacrée aux tests pédagométriques, nous avons communiqué les données susceptibles d'initialiser et d'encourager leur utilisation. Toutefois, ce n'est pas parce que le test représente un outil de mesure commode, pratique, rapide, souple, fiable,... qu'il faille en user et en abuser. Tout excès est nocif et porte la marque d'une détérioration virtuelle. Si l'usage des tests permet de progresser dans la résolution des problèmes que pose l'évaluation pédagogique, il est certain qu'il ne constitue pas une solution unique et universelle. BACHER souligne plusieurs objections à un usage généralisé et extensif des tests de connaissances sur la base des constats réalisés dans les pays où il en est ainsi.

Le test, tout comme un examen, présente un caractère ponctuel et il y a là un risque d'erreur important :
"... le niveau des élèves, apprécié en une seule occasion, ne peut l'être que de façon imparfaite. Chacun sait que ses propres performances varient d'une occasion à l'autre, sous l'effet de dispositions temporaires ("forme", fatigue, difficultés personnelles, etc.)".

Par rapport à cet aspect des choses, il faut souligner une fois encore la nécessité de relativiser tout résultat de toute évaluation. Rappelons qu'une bonne performance au test indique les capacités du sujet mais qu'une mauvaise performance ne permet pas de déterminer si la capacité est absente ou bien si elle ne s'est pas manifestée. On peut contrôler par comparaison avec les notes obtenues tout du long de l'année ou du stage, par un re-test avec une forme parallèle et si nécessaire aller plus loin dans une phase de diagnostic.

La forme et la nature souvent inhabituelles des épreuves, la complexité relative des consignes constituent des difficultés spécifiques aux tests. Ces difficultés nécessitent de la part des sujets la mobilisation de leurs capacités d'adaptation, surtout quand le test est donné en temps limité. Et les capacités d'adaptation n'ont rien à voir avec l'appréciation d'un niveau de connaissances. La familiarisation avec ce type d'épreuve améliore la performance. Un usage intensif des tests constituerait donc un nouveau contexte général et nécessiterait leur réétalonnage.

Nous avons déjà évoqué les problèmes des fraudes, du "bachotage", de la négligence des domaines d'expression et de créativité qui n'interviennent pas dans le test... Ces problèmes ne manquent pas de se manifester quand le test s'inscrit dans le cadre d'un enjeu social.

A cela, nous ajouterons un autre danger, celui de la catégorisation des individus par leur niveau au test, catégorisation définitive, absolue qui débouche sur des décisions sociales au caractère discriminatif. On a pu observer ce phénomène avec les tests d'intelligence utilisés pour alimenter les classes de l'enseignement spécial. Au nom d'un nombre devenu quasiment mythique, le Quotient intellectuel, on a exclu des milliers d'élèves du cycle scolaire normal. Il est à craindre qu'un usage administratif des résultats aux tests ne produisent les mêmes pernicieuses dérives. Et pourtant un bas niveau ne montre pas une incapacité. Dans une étude menée par BACHER et REUCHLIN et portant sur un groupe de 3 124 élèves du Cours moyen 2ème année suivis en sixième puis en cinquième, il est montré que la valeur prédictive des notes scolaires et des résultats à des tests d'intelligence et à des tests de connaissances obtenus au cours moyen reste faible quand on les compare aux classements établis ultérieurement. Plus finement, j'ai montré à l'occasion de la mise au point du *test prédictif pour le cours préparatoire* que si les enfants présentent des chances de réussir inégales en fonction de leur niveau initial, dans un rapport de 1 à 6, tous conservaient une probabilité de réussite. La valeur prédictive n'est jamais un absolu, seulement une probabilité. Quand cette probabilité, ce risque pédagogique, est connu alors le pédagogue peut agir en conséquence en étant attentif à chacun d'une manière différenciée en fonction du risque.

Dernier point, l'interprétation extensive du résulat obtenu au test de connaissances. Un test de niveau est un outil de classement, un point c'est tout. Il situe une performance, mais ne l'explique pas. Trop souvent on entend des interprétations bouclées, circulaires où les éléments sont tout à la fois cause et conséquence l'un de l'autre: tel élève est doué, pourquoi? parce qu'il a réussi le test; tel élève a réussi le test, pourquoi? parce qu'il est doué. Si on veut aller plus loin et dépasser le constat d'un niveau pour aller vers la compréhension des déterminants de ce niveau, il convient de relativiser le résultat obtenu au test.

Le meilleur moyen de relativiser un test, c'est de l'inclure dans une démarche évaluative plus globale qui fonctionne en tant que questionnement sur le sens, en tant que recherche de significations, en tant que construction d'une interprétation, en tant qu'élaboration d'une herméneutique. Pour cela, il faut toujours référer le test à un ensemble de contextes et aux relations qui unissent ces contextes.

Parmi ces contextes, le premier est la globalité de la personne. La performance est le produit d'un fonctionnement psychologique qui s'aborde comme processus fonctionnel mettant en jeu les opérations, les stratégies et les régulations adaptées à la tâche. Ce processus s'explore par des démarches diagnostiques.

Mais ce fonctionnement mental est lui-même déterminé par les circonstances d'une histoire individuelle dont il peut être utile de connaître grandes lignes et faits saillants. Plus que par le questionnaire administratif, c'est par l'établissement

d'une relation de confiance ente le formateur et le formé que ces déterminants pourront être saisis.

Un second contexte définit le groupe classe et la place du sujet dans ce groupe classe. Il est certain qu'une classe où règne le désordre et l'indiscipline aura des effets déstructurants sur les individus et que les sous-rendements y seront plus probables qu'ailleurs. La place sociale, selon que le sujet tient un rôle de "leader" ou un rôle de "bouc-émissaire", aura probablement une incidence sur la performance.

Un contexte plus large sera l'établissement, sa zone d'implantation géographique et de recrutement social. On sait la diversité des niveaux moyens d'un établissement à l'autre et selon la "cote sociale" de l'établissement, la performance s'interprètera différemment.

N'oublions pas dans cette évocation rapide et conclusive un fait essentiel et dont on n'a pas encore pris forcément toute la mesure: **la performance d'un élève ou d'un stagiaire est le produit direct d'une relation éducative**. On connaît bien cette attitude universelle qui fait que je m'attribue le mérite d'une réussite mais que je rejette les causes d'un échec sur les autres. Tel élève ou tel stagiaire réussit parce que je suis un pédagogue compétent, tel autre échoue parce qu'il présente un handicap quelconque. La recherche de sens dans l'interprétation évaluative passe par une forme d'auto-critique, d'auto-examen de conscience, d'auto-évaluation, d'auto-bilan de l'enseignant et du formateur. "Qu'ai-je fait qui a pu faciliter la réussite de tel ou tel, qu'ai-je fait ou que n'ai-je pas fait qui a pu empêcher une réussite?..." Bien entendu, cette réflexion peut se socialiser dans le cadre d'un travail d'équipe. Que l'on me comprenne bien! Je n'invite pas à cultiver les délices et les tortures morales de l'auto-culpabilisation. Les affinités interpersonnelles existent, se manifestent dans tous groupes sociaux, et il faut faire avec. Le manque de clairvoyance, de réflexes ou de réactions adaptées au moment voulu est un fait banal, commun à tout individu. Certaines difficultés ou erreurs de comportement peuvent objectivement être attribuées à l'élève ou au stagiaire... C'est beaucoup mieux quand on se donne les moyens de ces prises de conscience afin de dépasser le simple vécu qui globalise dans le lieu commun ou dans l'émotion le constat évaluatif. C'est beaucoup mieux parce qu'on peut commencer à comprendre les phénomènes sociaux et psychologiques qui génèrent un niveau de performances. On apprend ainsi à mieux connaître l'autre, à mieux se connaître, à percevoir le fonctionnement d'une relation éducative et à isoler ses éléments interactifs. Si je fais cela de telle manière avec telle personne et dans telles conditions j'obtiendrai probablement tel effet.

C'est par cette réflexion tournée vers l'analyse évaluative que le formateur peut progressivement élargir son répertoire de techniques instrumentales et de techniques relationnelles. Il pourra alors plus facilement atteindre les objectifs pédagogiques qu'il s'est définis ou que l'autorité lui a assignés.

La recherche de sens au moyen de l'analyse évaluative ne peut se contenter de présenter un aspect formel. Il ne suffit de comprendre pour comprendre, encore faut-il agir. Classer les élèves en situant leurs performances, comprendre leurs modes de fonctionnement grâce aux tests et aux démarches diagnostiques, saisir les effets de l'interaction et de l'influence sociales... n'auront d'utilité pratique

qu'à la condition que ces découvertes alimentent la création, l'entretien, la stimulation d'une dynamique d'apprentissage par l'évaluation. Cela veut dire concrètement que le formateur ne doit pas garder le produit de ses analyse pour lui tout seul. Il convient de les restituer aux apprenants de telle sorte que cela les aide à fixer leurs objectifs et à définir les moyens de leur atteinte. C'est probablement la phase la plus délicate. Rien de plus facile que de donner des résultats positifs. Mais lorsque les résultats sont mauvais, il faut trouver le moyen de les restituer sans provoquer de découragement, sans dévaloriser la personne, mais en quelle sorte comme un défi à surmonter. **En matière d'évaluation le respect de la personne passe avant toute chose.**

Par delà les aspects déontologiques qui accompagnent toute pratique sociale, le besoin d'outils, de démarches et de méthodes évaluatives fiables et efficaces se fera toujours sentir. La question posée est plutôt celle de leur adaptation aux finalités et aux objectifs de l'éducation et de la formation définis par les institutions. Il est évident que si l'objectif du système éducatif reste le tri social, l'évaluation formative, l'analyse étiologique des erreurs, la mobilisation des ressources pour entretenir les dynamismes de l'apprentissage qui font alterner la rigueur de la discipline et la jubilation que provoquent les maîtrises nouvellement acquises ne sont pas près de pouvoir s'ouvrir un champ d'expression et d'élaboration significatif.

Tout comme les méthodes pédagogiques, les méthodes d'évaluation se présentent comme fondamentalement contradictoires puisqu'elles doivent participer à la formation du citoyen de demain, alors qu'elles restent coalescentes des mentalités d'aujourd'hui. C'est en ouvrant l'espace contradictoire qu'elles trouveront les ressources de leurs évolutions.

TABLE DES PLANCHES

GLOSSAIRE DES SIGLES UTILISES

- A.F.P.A. : association nationale pour la formation professionnelle des adultes
- B.B.C. : british broadcasting corporation
- B.E.P. : brevet d'enseignement professionnel
- C.A.P.U.C. : certificat d'aptitude professionnelle par unités capitalisables
- C.E.1. : cours élémentaire première année
- C.E.P. : certificat d'études primaires
- C.F.A. : centre de formation des apprentis
- C.F.G. : certificat de formation générale
- C.P. : cours préparatoire
- C.P.A. : centre de psychologie appliquée
- C.R.D.P. : centre régional de documentation pédagogique
- S.R.E.S.A.S. : section de recherche de l'éducation spécialisée et de l'adaptation scolaire
- E.A.P. : établissements d'applications psychotechniques
- G.F.E.N. : groupe français d'éducation nouvelle
- I.N.R.A.P. : institut national de recherches et d'applications pédagogiques
- I.N.R.P. : institut national de la recherche pédagogique
- I.N.S.E.E. : institut national des statistiques et des études économiques.
- L.E.P. : lycée d'enseignement professionnel
- O.N.I.S.E.P. : office national d'information sur les enseignements et les professions.
- O.T.I. : objectif terminal global ou d'intégration
- Q.C.M. : questionnaire à choix multiples
- T.M.C.P. – T. : test de mathématiques pour le cours préparatoire – test.
- T.P.6. – T. : test pour le passage en sixième – test.

ANNEXE

LISTE DES TESTS PEDAGOMETRIQUES

Les différentes épreuves sont présentées en trois parties. Le domaine scolaire, formation générale (domaine A) regroupe les tests utilisables depuis l'école maternelle. Le domaine de la formation professionnelle et technique initiale (domaine B) reprend les épreuves étalonnées pour des élèves fréquentant les établissements qui préparent à un métier. Le dernier domaine, celui de l'entreprise et de la formation continue (domaine C) rassemble les tests réalisés pour des publics ayant quitter les écoles et les universités.

La présentation se fait sur dix colonnes :

- colonne n° 1: titre du test et nom de l'auteur; ce nom peut être celui d'une personne ou celui d'une institution.

- colonne n° 2: objet du test; on note ici à quel usage est destiné le test.

- colonne n° 3: étalonnage; apparaît à cet endroit la caractéristique essentielle qui définit le groupe de référence et donc le type de population auquel le test est applicable.

- colonne n° 4: E comme E.A.P.; un "x" dans cette colonne indique que le test est édité et/ou diffusé par les E.A.P..

- colonne n° 5: C comme C.P.A.; un "x" dans cette colonne indique que le test est édité et/ou diffusé par les C.P.A..

- colonne n° 6: I pour individuel; un "X" dans cette colonne indique que la passation est individuelle par construction; un "x" veut dire que l'épreuve est de passation collective et qu'elle se prête également à une passation individuelle.

- colonne n° 7: C pour collectif; un "X" dans cette colonne est l'indication d'une passation collective.

- colonne n° 8: N pour niveau; un "X" dans cette colonne indique que le résultat du test situe un niveau de connaissances.

- colonne n° 9: P pour pronostic; un "X" dans cette colonne indique que l'épreuve donne une prédiction ou une probabilité de réussite dans un apprentissage par construction; un "x" indique que la prédiction n'est pas l'objectif premier du test.

- colonne n° 10: D pour diagnostic; un "X" indique que le test a été conçu pour mener un diagnostic; un "x" indique que la forme des résultats peut être utile pour amorcer une démarche de diagnostic.

Cette annexe a été conçue dans le but d'aider à dégrossir le choix d'un test adapté aux besoins de l'utilisateur. On peut opérer une première sélection d'épreuves à l'aide du tableau en faisant jouer successivement chaque critère que l'on aura défini par rapport à son besoin. Par exemple, on fait ressortir toutes les épreuves prédictives pour le C.P. Puis on ne garde que celles qui ont une passation collective. Pour déterminer le choix définitif parmi les épreuves ainsi sélectionnées, on pourra se reporter au catalogue de l'éditeur pour comparer les données relatives aux qualités métriques respectives et enfin, pour lever un dernier doute on termine par la comparaison des specimen-set dont on aura passé la commande.

Dans cette liste, nous avons fait figurer toutes les épreuves rangées dans les rubriques "tests de connaissances" par les éditeurs. Nous avons repris également quelques "tests cliniques" et "tests d'aptitudes" lorsqu'ils présentent un rapport direct avec une démarche diagnostique à caractère pédagogique.

A - DOMAINE SCOLAIRE, FORMATION GENERALE

A.1 : maternelle et primaire

A.1.1 - grande section de maternelle et cours préparatoire: épreuves à caractère prédictif

1 Titre / auteur	2 Objet du test	3 Etalonnage	4 E	5 C	6 I	7 C	8 N	9 P	10 D
Test prédictif pour cours préparatoire - T.P.C.P. - Pasquier	Probabilité de réussite en lecture au C.P.	Début C.P. redoublants sexe et milieu	x		x	X		X	x
Echelle d'admission en cycle élémentaire - E.A.C.E. - Tettelin - Chautard	Prédiction de l'adaptation globale aux apprentissages	Enfants de 5 à 7 ans	x			X		X	x
Batterie de lecture et prédictive - INZ - Inizan	Prédiction de l'opportunité et de la durée de l'apprentissage de la lecture Niveau de lecture	Enfants de 5 à 7 ans	x		X X	x	X	X	
Test des concepts de base - BOEHM - Boehm	Mesure de la maîtrise des concepts nécessaires à la réussite	Début de grande section et du C.P.		x	x	X			x
Test reversal - Y5 - Edfelt	Prédiction de la réussite en lecture	Fin de G.S. début du C.P.		x	x	X		X	
Echelle d'évaluation et batterie analytique - N.B.A. - Ravard - Rabreau	Prévention de l'échec scolaire, pédagogie par objectifs	G.S. C.P.	x			X	X	X	x
Contrôle des aptitudes à la lecture et à l'écriture - C.A.L.E. - Girolami - Boulinier	Vérification des aptitudes		x						X

A.1.2 – primaire (du C.P. au C.M.2)

1 Titre / auteur	2 Objet du test	3 **Etalonnage**	4 E	5 C	6 I	7 C	8 N	9 P	10 D
Test de lecture pour cours préparatoire - T.L.C.P. - Pasquier	Bilan des acquis en lecture pour assurer la tran - sition avec le CE	Fin C.P.	x		x	X	X	x 1	x
Test de mathématiques pour cours prépara - toire - T.M.C.P. - Pasquier - Thébault	Bilan des acquis en mathématiques, analyse des erreurs	Fin C.P.	x		x	X	X		X
Evaluation du savoir - lire - LIRE - Inizan	Evaluation de la compétence en lecture	Enfants du C.P.; mars et juin	x		x	X	X		x
Test de connaissances: cours préparatoire - C.P.C.M. - Cuiller - Maury	Connaissances en français et en calcul	Fin C.P. début C.E.1 garçons filles redoublants	x		x	X	X		x
Test de français pour cours élémentaire - T.F.C.E. - Pasquier - Chamaulte	Bilan complet des connaissances, esti - mation de niveau, analyse des erreurs	fin C.E.2.	x		x	X	X	x	x
Test de mathématiques pour cours élémentaire - T.M.C.E. - Pasquier	Bilan complet des connaissances, analyse des erreurs, pronostic	Fin C.E.2	x		x	X	X	x	x
Test de passage en sixième - T.P.6. - Pasquier	Probabilité de passage en 6è niveau fin de cycle primaire	Fin de C.M.2	x		x	X	X	x 2	x
Echelle d'évaluation mathématiques - NBM10 - NBM11 - NBM20 - NBM21 - NBM30 - NBM31 - NBM40 - NBM41 Ravard	Contrôle des connais - sances en calcul Soutien orientation en sixième	 Fin C.E.1 Fin C.E.2 début C.M.1 fin 2è trim. sébut C.M.2 fin 2è trim.	x		x	X	X		x

A.1.2 – suite

1 Titre / auteur	2 Objet du test	3 Etalonnage	4 E	5 C	6 I	7 C	8 N	9 P	10 D
Test de niveau mathématiques – AME – Aufaure	Niveau des acquisi - tions, recherche des difficultés d'assimilation	Fin C.P., CE1, CE2, CM1, CM2	x		x	X	X		x
Evaluation et analyse des acquisitions scolaires – BATLM – Savigny	Echelles d'apprentis - sage en lecture, calcul et orthographe	Classes élémen - taires	x		X		X		
– RUP –	Analyse des erreurs orthographiques	7 à 12 ans	x		x	X			X
Echelle de quotient scolaire – E.Q.S.64 – Bourdier	Déterminer le niveau scolaire, calcul d'un quotient scolaire	Fin d'année du CP au CM2 6 à 11 ans	x		X		X		
Tests de niveau scolaire de Subes, orthographe et calcul – Y1 – Subes – Subes	Déterminer le niveau scolaire pour répartir les élèves dans les classes	Classes primaires garçons filles		x	x	X	X		
Test du cycle élémentaire – T.C.E. – Lepez – Riquier	Evaluation du niveau scolaire en français et en mathématiques	Elèves du CE1 au CM2		x	x	X	X	x	
Bilan de calcul – MAT.MA. – Rochettes	Epreuve pour les enfants de l'école primaire bloqués en calcul; bilan des acquis cognitifs et des lacunes; appréciation de la maturité mathématique.		x		X				X

Notes :

1 – Les données relatives aux probabilités de passage au C.E.1 puis au C.E.2 établies à partir des résultats obtenus au T.L.C.P. figurent à la page 15 du manuel du T.P.C.P.

2 – On trouvera les données du suivi des élèves établies pour les classes de sixième et de cinquième des collèges dans *Psychologie et psychométrie*, 1988, 9, 2, pp. 51 – 54.

A.2 – Primaire et au-delà

1 Titre / auteur	2 Objet du test	3 Etalonnage	4 E	5 C	6 I	7 C	8 N	9 P	10 D
Batterie orthographe - lecture - ORLEC - Beaumont/Oise	Etude analytique de tous les aspects de la lecture et de l'orthographe.	Fin CP au CM2 et entrée en 6ème	x						X
Test d'évaluation du savoir lire: "la pipe et le rat" - LIREL - Lefavrais	Niveau de lecture, rapidité de compré - hension, coefficient d'éducabilité	Du CP au lycée	x			X	X		x
L'ALOUETTE Lefavrais	Analyse de la lecture, diagnostic de la dys - lexie, âge de lecture	Enfants du primaire, collèges, lycées, adultes		x	X		X		X
Test de niveau d'orthographe - T.N.O. - Doutriaux - Lepez	Evaluation du niveau d'orthographe	CE2 à terminale adultes		x	x	X	X		
Tests d'acquisitions scolaires - T.A.S. - Lepez - Benat - Lepez - Cahen - Gleize - Kourovsky - Lemaitre	Mesure des niveaux de connaissances en français, mathéma - tiques, anglais, allemand	CE1 - CE2 CE2 - CM1 CM1 - CM2 CM2 - 6è 6è - 5 5è - 4è 3è - 2è fin 1re		x	x	X	X		
Fiches calcul et français - Y.C.F. - I.N.E.T.O.P.	Niveau scolaire	Elèves de CM1 et 2 FEP	x		x	X	X		
Epreuves standardisées de raisonnement mathématique - Y.R.M. - I.N.E.T.O.P.	Niveau logico - mathématique	Début CE2, début CM1 début CM2, fin CM2 début 6è, fin 6è début 5è, début 4è début 3è, début 2de	x		x	X	X		
Test de lecture silencieuse - Y.L.S. - I.N.E.T.O.P.	Compréhension de la la lecture silencieuse	Début CE2 ... 3è	x		x	X	X		

A.2 – suite

1 Titre / auteur	2 Objet du test	3 Etalonnage	4 E	5 C	6 I	7 C	8 N	9 P	10 D
Test analytique de mathématiques – Y.T.A.M. – I.N.E.T.O.P.	Lacunes et/ou difficultés du développement de la pensée mathématique du CM2 à la 5è	Elèves de 6è faibles 6è normales 5è faibles	x		X	x	x		X
Test de compréhension de concepts numériques – T.C.N. – Simon	Niveau atteint dans la compréhension du sens des opérations, analyse des difficultés	10 ans 11 ans 12 ans 13 ans		x	x	X	X		x
Test d'évaluation des capacités métacognitives – T.E.C.M. – Pasquier	Evaluation analytique de dix capacités métacognitives intervenant dans les apprentissages et l'adaptation, prescription d'un programme de remédiation cognitive pour tout sujet sachant lire et écrire		x		x	X			X 3

Notes :

3 – Des stages de formation à l'utilisation du T.E.C.M. sont organisés par les E.A.P.

A.3 – Collèges et lycées

1 Titre / auteur	2 Objet du test	3 Etalonnage	4 E	5 C	6 I	7 C	8 N	9 P	10 D
Connaissances en français – MAFI – Visieux – MAFO – – MAFU – Mounaix	Evaluation des connaissances générales en français	1è et 2de 2de et 3è 3è et 4è	x		x	X	X		
Connaissances en anglais – MASA – Hays	Estimation des connaissances requises	fin de 1è début de terminale	x		x	X	X		

A.3 – suite

1 Titre / auteur	2 Objet du test	3 Etalonnage	4 E	5 C	6 I	7 C	8 N	9 P	10 D
Test de connaissance en anglais – Y.C.AN – I.N.E.T.O.P.	Niveau en anglais	3è	x		x	X	X		
Connaissances début sixième – Y.C.D.6. – I.N.E.T.O.P.	Bilan des connaissances en français et en calcul	Début 6è	x		x	X	X		
Calculs et problèmes – Y.C.P.5. – I.N.E.T.O.P.	Sondage des connaissances de base en mathématiques	5è au mois d'avril	x		x	X	X		
Epreuve de connaissances en mathématiques – Y.E.C.M.3. – I.N.E.T.O.P.	Niveau	3è, fin du second trimestre scolaire	x		x	X	X		
Inventaire de connaissances en français – Y.I.C.S.1. – I.N.E.T.O.P.	Connaissance de la littérature française	Début d'année, 1è, 2è modernes classiques	x		x	X	X		
Inventaire de connaissances en français et en calcul – Y.I.C.S.2 à 6 – I.N.E.T.O.P.	Bilan de connaissances aux périodes de changement de classes	1er cycle secondaire	x		x	X	X		
Test de connaissances littéraires, – Y.T.C.L. – Test de connaissances mathématiques – Y.T.C.M. – I.N.E.T.O.P.	Niveau	fin de 3è	x		x	X	X		

B – ETABLISSEMENTS PREPARANT A UN METIER

1 Titre / auteur	2 Objet du test	3 Etalonnage	4 E	5 C	6 I	7 C	8 N	9 P	10 D
Connaissances calcul et français - Y.C.F. 4 et 6 - I.N.E.T.O.P.	Examens d'entrée dans les C.E.T. (actuellement les L.E.P.)	Garçons filles	x		x	X	X		
Connaissance en électricité - CEA -	Niveau	C.A.P. d'électriciens	x		x	X	X		

C – ADULTES ET FORMATION CONTINUE

1 Titre / auteur	2 Objet du test	3 Etalonnage	4 E	5 C	6 I	7 C	8 N	9 P	10 D
Test de connaissances mécaniques - D.C.T.A. - Alquier	Test de discrimination de connaissances techniques	Fins d'années d'école professionnelle	x		x	X	X		
Connaissances en dessin industriel - C.D.I. - Morteveille	Niveau	18 à 50 ans	x		x	X	X		
Connaissances en mathématiques - MAT - Xydias	Déterminer un niveau hiérarchique	Adultes manoeuvres à cadres	x		x	X	X		
Connaissances en hydraulique - CHF - Formilor	Niveau	Adultes	x		x	X	X		
Connaissances juridiques - CJLG - Leparmentier	Inventaire des connaissances que doit savoir normalement tout licencié en droit		x		x	X			

BIBLIOGRAPHIE

La bibliographie sur l'évaluation pédagogique et sur les tests représente un ensemble conséquent. Seuls sont répertoriés ici les titres et leurs auteurs cités en référence dans l'ouvrage.

- BACHER (F.) La docimologie, in ***Traité de psychologie appliquée, 6, Education et institutions éducatives,*** Paris, P.U.F., 1972, pp. 25 - 87.

- BACHER (F.) et REUCHLIN (M.) Le cycle d'observation. Enquête sur l'ensemble des élèves d'un département, ***B.I.N.O.P.,*** 1965, 21, pp. 149 - 236.

- BALLION (M.) Signalements de difficultés et devenir scolaire, ***S.R.E.S.A.S.,*** 1979, 20.

- BARBIER (J.M.) ***L'évaluation en formation,*** Paris, P.U.F., 1985.

- BEAUVOIS (J.L.) et ENRIQUEZ (E.) Editorial, ***Connexions,*** 1976, 19, pp. 3 - 4.

- BINOIS (R.) ***La psychologie appliquée,*** 8ème édition, Paris, P.U.F., 1965.

- BONBOIR (A.) ***La docimologie,*** Paris, P.U.F., 1972.

- BONNIOL (J.J.) Recherches et formations: pour une problématique de l'évaluation formative, in ***L'évaluation: approche descriptive ou prescriptive?,*** Bruxelles, De Boeck - Wesmael, 1986, pp. 119 - 133.

- BOREL - MAISONNY (S.) ***Langage oral et écrit,*** Neuchâtel, Delachaux et Niestlé, 1960.

- BOURDIEU (P.) et PASSERON (J.C.) ***La reproduction,*** Paris, Minuit, 1970.

- BLOOM (B.S.) et col. ***Taxonomie des objectifs pédagogiques,*** trad. franc., Montréal, Education nouvelle, 1969.

- BUJAS (Z.) L'influence du mode de notation sur la validité des tests de connaissances, ***Travail humain,*** 1965, 28, pp. 197 - 202.

- CARDINET (J.) ***Evaluation scolaire et pratique,*** Bruxelles, De Boeck - Wesmael, Neuchâtel, I.R.D.P., 1986.

- CARDINET (J.) ***Pour apprécier le travail des élèves,*** Bruxelles, De Boeck - Wesmael, 1986.

- C.R.D.P. de LYON ***Docimologie et examens,*** Lyon, I.P.N., 1969.

- CUSIN et PIOLAT Auto - évaluation **a posteriori.** Quelques déterminants scolaires et intellectuels, ***Cahiers de psychologie,*** 1972, 15.

- DEMANGEON (M.) et LARCEBEAU (S.) Une expérience de correction multiple, ***B.I.N.O.P.,*** 1958, 14, 131 - 156.

- DOMINICE (P.) ***La formation, enjeu de l'évaluation,*** Berne, Peter Lang, 1979.

- DUSSAULT (G.) Sept cents instruments de SAGE - français pour l'évaluation des travaux et des apprentissages en langue maternelle à l'école primaire: types et caractéristiques, in ***Actes du 27ème colloque de l'A.I.P.E.L.F.,*** Lausanne, Centre vaudois de recherches pédagogiques, 1983.

- FERRE (A.) ***Les tests à l'école,*** 8ème édition, Paris, Bourrelier - Colin, 1968.

- FEUERSTEIN (R.) et col. ***Instrumental enrichment,*** Baltimore, University Park Press, 1980.

- GILLY (M.) L'élève vu par le maître: influences socio - normatives dans l'exercice du rôle professionnel, in ***L'évaluation: approche descriptive ou prescriptive?,*** De Boeck - Wesmael, 1986, pp. 71 - 90.

- GOGUELIN (P.) ***Psychologie industrielle,*** Paris, C.N.A.M.

- GUIGOU (J.) ***Critique des systèmes de formation,*** Paris, Anthropos, 1972.

- GUILBERT (J.J.) Opinion des étudiants sur le système des examens par Q.C.M. après une année d'expérience, ***Revue du praticien,*** suppl., 1962, 12, 18, pp. I - V.

- IFAPLAN ***Transition of young people from education to adult and working life,*** Nouvelles tendances de l'évaluation pédagogique: Les profils, DOC: 22 WD 84 FR, Bruxelles, Programme information office, IFAPLAN, 1984.

- JACQUARD (A.) in PINEL (D.) Une technique ou une culture? ***Le Monde dossiers & documents,*** avril 1989.

- KETELE (J.M. de) et PAQUAY (L.) Pédagogie corrective, énoncé des objectifs pédagogiques et évaluation formative, ***Sciences de l'éducation,*** 1974, 1, pp. 85 - 96.

- KOCHER (F.) ***La rééducation des dyslexiques,*** Paris, P.U.F., 1970.

- LANDSHEERE (G. de) ***Evaluation continue et examens. Précis de docimologie,*** Bruxelles, Labor, Paris, Nathan, 1971.

- LANDSHEERE (G. de) ***Dictionnaire de l'évaluation et de la recherche en éducation,*** Paris, P.U.F., 1979.

- LAUGIER (H.) et WEINBERG (D.) Elaboration statistique des données numériques de l'enquête sur la correction des épreuves du Baccalauréat, in ***La correction des épreuves écrites dans les examens,*** Paris, Maison du livre, 1936.

- LAUGIER (H.), PIERON (H.), PIERON (Mme H.), TOULOUSE (E.), WEINBERG (D.), Etudes docimologiques sur le perfectionnement des examens et concours, ***Publications du Travail humain,*** A, 3, 1934.

- LEFAVRAIS (P.) Quelle utilisation du test de l'alouette, ***Le journal des psychologues,*** octobre 1989, 71, pp. 45 - 49.

- LEPLAT (J.) et PETIT (R.), Un aspect pédagogique du principe de la connaissance des résultats ***Bulletin du CERP,*** 1962, XI, 4.

- LOBROT (M.) Une doctrine: contre les examens, in ***Pourquoi des examens?,*** Paris, Société des éditions rationalistes, 1968.

- LURCAT (L.) ***L'échec et le désintérêt scolaire à l'école primaire,*** Paris, C.E.R.F., 1976.

- LURCAT (L.) ***Une école maternelle,*** Paris, Stock, 1976.

- MEIGNIEZ (R.) et VALATX (M.C.) L'évaluation en formation des adultes, ***Pour,*** 1972, 20.

- NOIZET (G.) et CAVERNI (J.P.) ***Psychologie de l'évaluation scolaire,*** Paris, P.U.F., 1976.

- NUTTIN (J.) ***Tâche, réussite et échec. Théorie de la conduite humaine,*** 3ème édition, Louvain, Publications universitaires, 1971.

- PASSERON (J.C.) Sociologie des examens, in ***Education et gestion,*** 1970, 20.

- PIAGET (J.) et INHELDER (B.) Les opérations intellectuelles et leur développement, in ***Traité de psychologie expérimentale, VII, L'intelligence,*** Paris, P.U.F., 1963, pp. 109 - 155.

- PIERON (H.), REUCHLIN (M.), BACHER (F.), DEMANGEON (M.), Analyse des corrélations entre notations à une session de baccalauréat, ***Biotypologie,*** 1962, 23, 1 - 2, pp. 17 - 47.

- PIERON (H.), REUCHLIN (M.), BACHER (F.), DEMANGEON (M.), ***Examens et docimologie,*** Paris, P.U.F., 1963.

- PIERON (H.), REUCHLIN (M.), BACHER (F.), DEMANGEON (M.), ***Vocabulaire de la psychologie,*** 6ème édition, Paris, P.U.F., 1979.

- PIERON (H.), REUCHLIN (M.), BACHER (F.), Une recherche expérimentale de docimologie sur les examens oraux de physique au niveau du baccalauréat de mathématiques, ***Biotypologie,*** 1962, 23, 1 - 2, pp. 48 - 73.

- POSTIC (M.) ***La relation éducative,*** Paris, P.U.F., 1979.

- REMONDINO (C.) Etude factorielle sur la notation de compositions scolaires portant sur la langue maternelle, ***Le travail humain,*** 1959, 22, pp. 27 – 40.

- REUCHLIN (M.) Problèmes d'évaluation, in ***Traité des sciences pédagogiques, 4, Psychologie de l'éducation,*** P.U.F., 1974, pp. 207 – 236.

- ROUSSET (D.) ***Cadres et formation,*** Bordeaux, Echelon régional de l'emploi, Ministère du travail, A.N.P.E., 1974.

- SCRIVEN (M.) The methodology of evaluation, in Perspectives on curriculum evaluation, ***AERA monograph series on curriculum evaluation,*** 1967, 1, Chicago, Rand Mac Nally.

- TOURNEUR (Y.) L'analyse d'objectifs dans l'évaluation de maîtrise. Etude de quelques indices, ***Sciences de l'éducation,*** 1974, 1, pp. 36 – 56.

PLANCHES

Certificat de Premier Cycle : Evaluation d'Elève

Elève : ..... Etablissement : .....

Période d'évaluation : .....

I. — COMMUNICATION

(Echelle d'évaluation : 1 à 10)

		Commentaires
Lecture		.....
Présentation écrite		.....
Expression orale		.....
Maniement des chiffres		.....
Utilisation de graphiques, cartes, etc...		.....

II. — DEVELOPPEMENT PERSONNEL ET SOCIAL

(Echelle d'évaluation : 1 à 10)

		Commentaires
Travail en commun		.....
Travail autonome		.....
Ponctualité		.....
Fiabilité		.....
Persévérance		.....
Initiative		.....

III. — APTITUDES PRATIQUES

(Echelle d'évaluation : 1 à 10)

		Commentaires
Ecoute, Exécution d'instructions ..		.....
Maniement d'équipement		.....
Coordination physique		.....

NOTE : La notation s'effectue de 1 à 10 ; la note la plus élevée (10) indique une maîtrise satisfaisante de la compétence concernée.

Tiré de IFAPLAN

Planche n° 1 : PROFIL PEDAGOGIQUE

Projet pilote des C.E. :

projet pour les jeunes quittant prématurément l'école, Irlande

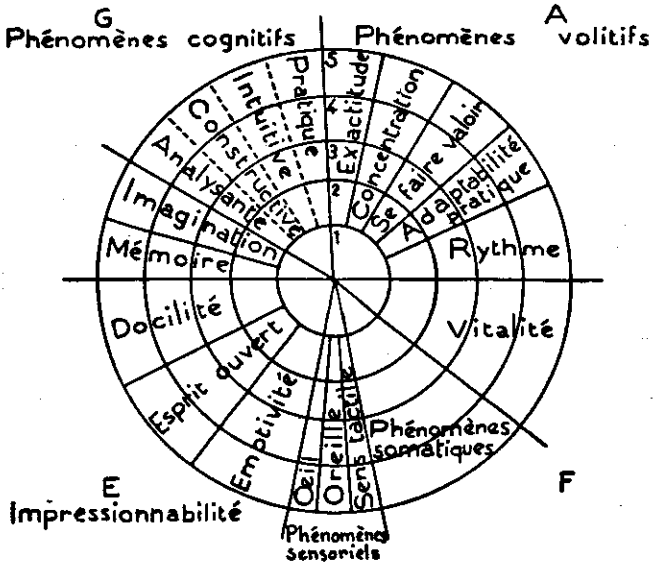
LE SYSTEME DU PROFIL D'ELEVE

ECHELLE DES NOTES : une explication de la note la plus élevée (10) pour chaque compétence est fournie ci-dessous.

COMMUNICATION	Note maximum (10).
Lecture	Peut comprendre un texte écrit si les idées y sont exprimées en termes simples.
Présentation écrite	Peut écrire un rapport, une lettre, etc..., clairs et précis.
Expression orale	Peut décrire des événements avec aisance, assurance et précision.
Maniement des chiffres....	Peut effectuer des calculs familiers ou simples.
Utilisation de graphiques, cartes, etc...	Peut interpréter et utiliser des croquis, des plans, etc...
CAPACITES PERSONNELLES ET SOCIALES	Note maximum (10).
Travail en commun	Fournit toujours une contribution positive aux activités du groupe.
Travail autonome	Se concentre toujours sur une tâche, garantissant personnellement un travail bien fait.
Ponctualité	Veille toujours à être à l'heure aux rendez-vous.
Fiabilité	Assume toujours la responsabilité de mener à bien les tâches allouées.
Persévérance	Fait toujours preuve de détermination pour mener les tâches à bien.
Initiative	Fait toujours preuve d'énergie et de ressources pour agir sans incitation ni instructions.
APTITUDES PRATIQUES	Note maximum (10).
Ecoute, exécution d'instructions	Fait toujours preuve d'attention et exécute les instructions avec compétence.
Maniement d'équipement ..	Sait manipuler avec compétence les outils et l'équipement.
Coordination physique	Maîtrise une grande variété de mouvements.

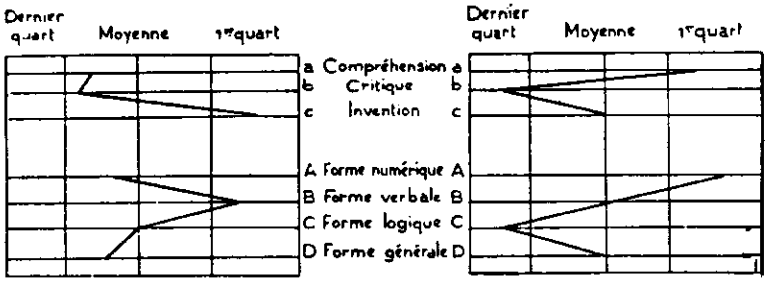
DUPONT René pour ajusteur outilleur	POURCENTAGE DE CHANCE POUR QUE L'INDIVIDU SOIT "BON"						
	(2) 0 % (S)	20 % (R)	40 % (Q)	(P) 60 % (O)	80 % (N)	100 % (M)	
test A1 représentant le critère A	9	19	23	26	30	40	44
test B1 représentant le critère B	15	18	19	20	22	24	26
test C1 représentant le critère C	7	15	18	20	23	26	27
test D1 représentant le critère D	30	35	39	43	48	49	55
test E1 représentant le critère E	2	15	17	18	21	22	33
test F1 représentant le critère F	3	7	12	15	17	20	25
test F2 représentant le critère F	23	37	45	50	57	60	70
test G1 représentant le critère G	16	23	25	26	29	34	35
test H1 représentant le critère H	----------						24 29
(test du type de la figure 6)							

Tiré de GOGUELIN

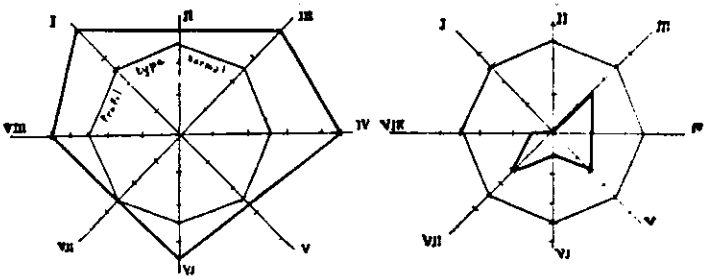


Cadre de Von MULHER (d'après FERRE)

Planche n° 2 : PROFILS PSYCHOMETRIQUES



Profil utilisé par PIERON pour des épreuves d'intelligence (d'après FERRE)



Profil utilisé par VIOLET-CONIL et CANIVET pour le BINET-SIMON (d'après FERRE)

I. - Compréhension et raisonnement abstrait.

II. - Jugement et esprit critique.

III. - Compréhension et raisonnement concret.

IV. - Connaissances acquises.

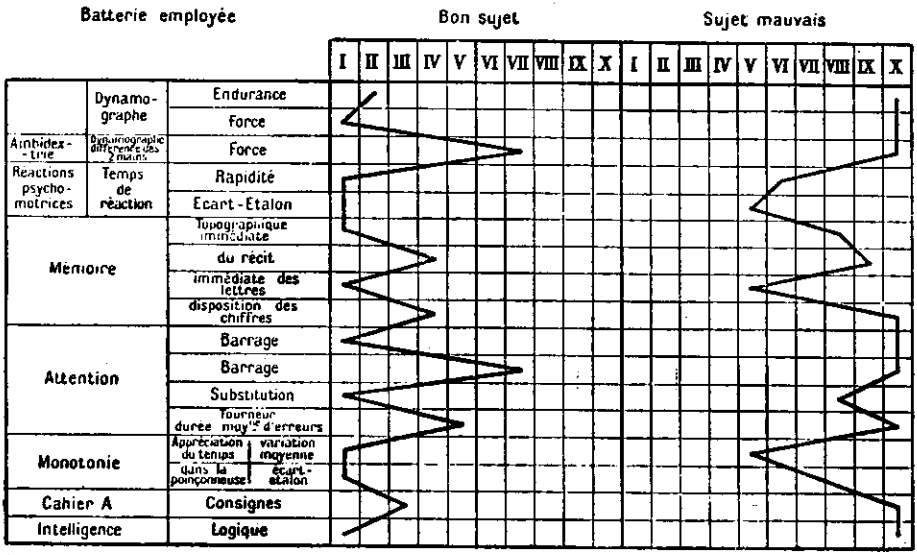
V. - Organisation des connaissances.

VI. - Mémoire.

VII. - Représentation spatiale.

VIII. - Synthèse.

Profil-type d'après LAHY et KORNGDD (tiré de BINOIS)



TMCP-E

NOM .....

Prénom .....

Notes Stand	Numération N (1 + 2 + 3)	Opérations O (4 + 5)	Problèmes P (6)	Logique L (7 + 8 + 9)	Espace E (10 + 11)	Notes Stand
14	30	8	5		18	14
13	28-29	7		18	17	13
12	25-27	6	4	16-17	16	12
11	22-24	5		15	14-15	11
10	19-21	4	3	13-14	13	10
9	17-18	3		12	12	9
8	14-16	2	2	10-11	11	8
7	11-13	1		9	10	7
6	9-10	0	1	7-8	9	6
5	6-8			6	8	5
4	3-5		0	4-5	7	4
3	1-2			3	6	3
2	0			1-2	4-5	2
1				0	3	1
0					0-2	0

Epreuves	Notes Stand
N	11
O	12
P	14
L	10
E	8
Total	55

Classement :

Inter déciles	1er	2è	3è	4è	5è	6è	7è	8è	9è	10è
points	68-64	63-61	60-58	57-55	54-52	51-50	49-46	45-42	41-34	33-13

Dépôt légal, 1er trimestre 1982

Planche n° 3 : PROFIL PEDAGOMETRIQUE

TP6M–T

CALCUL

ÉPREUVE DE CALCUL

Nom : ..... *Classe* : .....

Prénom : ..... *Date* : .....

– CONSIGNES –

Ce test de mathématiques comprend 6 parties : numération, opérations, mesures, problèmes, fonctions numériques, géométrie.

Lis bien attentivement les consignes suivantes qui t'expliquent comment procéder.

Pour répondre aux questions posées, il te suffira de placer une croix et une seule en face de la bonne réponse.

Voici un exemple :

	1,8	
18 : 1 000	18 000	
Le résultat de cette opération est	0,018	X
	9	

Tu remarques qu'on a tracé une croix en face de la bonne réponse 0,018. Il n'y a qu'une seule croix à placer car il n'y aura toujours qu'une seule bonne réponse à chaque question.

Tu remarques également qu'à gauche de la dernière case il n'y a rien d'écrit sur la ligne.

C'est fait exprès. Si tu penses que la bonne réponse n'est pas parmi les réponses proposées tu pourras ainsi disposer d'une ligne pour inscrire la réponse que tu auras trouvée.

Si tu n'as pas bien compris ces explications, tu demandes tout de suite au maître ou à la maîtresse de te dire ce qu'il faut faire.

Planche n° 4 : CONSIGNE DE Q.C.M.

Tu feras bien attention : mettre une croix c'est vite fait. Mais il faut qu'elle soit en face de la bonne réponse. Alors réfléchis bien avant de noter ta réponse.

Si une question te paraît trop difficile, tu passes à la suivante pour ne pas perdre trop de temps.

Si tu as besoin de poser une opération, ou de faire un petit dessin, un tableau pour trouver ta réponse tu utiliseras la partie notée (brouillon) et située à gauche de la question. Quand il n'y a pas de partie (brouillon) cela veut dire que tu dois trouver directement la réponse.

Tu travailles à ta vitesse, et tu refermeras le cahier dès que tu auras terminé.

Attends le signal avant de commencer.

TESTS DE NIVEAU SCOLAIRE DE SUBES ORTHOGRAPHE ET CALCUL

J. et M. SUBES

Centre de Psychologie Appliquée

Matériel

Cahier de Test : *12 pages.*
Manuel : *24 pages.*

Administration

Passation : *individuelle ou collective.*
Temps de passation : *40 minutes à 2 heures suivant le niveau.*
Temps de correction : *2 à 5 minutes.*
Age d'application : *classes primaires.*
Indications : *détermination du niveau scolaire.*
Etalonnage : *enfants français.*

NOTICE TECHNIQUE

Ces *tests de niveau scolaire* ont été construits pour permettre la répartition des élèves dans les différentes classes et sections des écoles primaires depuis le Cours Élémentaire jusqu'à la classe de Fin d'Études Primaires.

Ils portent sur l'Orthographe et le Calcul et sont divisés en cinq subtests qui correspondent aux niveaux des différentes classes. Chaque subtest comprend, pour l'orthographe, cinq phrases dont les mots et les tournures grammaticales ont été choisis pour s'adapter aux possibilités normales des élèves de chacune des classes, et pour le calcul, 7 à 8 exercices ou problèmes simples. Pour examiner un enfant il suffit d'avoir recours aux subtests qui avoisinent son niveau probable.

Cette échelle de tests tient compte des programmes officiels de l'enseignement primaire et recouvre tous les cours et l'ensemble des connaissances en orthographe et en calcul. L'échelle est graduée en classes et non en âges, en fonction du niveau des élèves au début de l'année scolaire.

Pour chaque subtest on dispose d'un étalonnage pour les filles, d'un étalonnage pour les garçons et d'un étalonnage d'ensemble. De plus, on a déterminé le nombre de fautes qu'un élève ne doit normalement pas dépasser pour être admis dans la classe correspondant à l'épreuve; c'est cette donnée qui permet la répartition des élèves dans les classes d'après leur niveau scolaire. Ces indications sont fournies d'une part pour les épreuves de calcul et les épreuves d'orthographe séparément et d'autre part pour l'ensemble de ces deux notes afin de permettre une appréciation globale du niveau scolaire de l'enfant.

Ces épreuves sont utilisables même par des éducateurs non familiarisés avec la pratique psychotechnique. Elles peuvent être utilisées soit sous forme collective soit sous forme individuelle.

Une étude portant sur la validité des épreuves et les conditions d'étalonnage a paru dans la *Revue de Psychologie Appliquée*, 2 (2), Avril 1952.

Rubrique	Sigle	TEST DE CONNAISSANCES MECANIQUES
6 2	DCTA	*M. ALQUIER*

Description :

Test : DCTA-T.
Feuilles de réponses : DCTA-F.
Grille de correction : DCTA-G.
Manuel : DCTA-C.
Spécimen-Set : DCTA-S.

Utilisation :

Passation : individuelle ou collective.
Temps de passation : 30 minutes.
Temps de correction : 2 minutes.
Age d'application : adolescents, adultes.

Statistiques :

Etalonnages : élèves sortant d'une école de formation professionnelle, tout-venant, professionnels d'atelier.

Présentation :

Ce test a été conçu pour évaluer le niveau de connaissances techniques en mécanique des candidats à l'embauche dans les professions qualifiées de la mécanique générale, ainsi que pour la sélection de candidats à la promotion ouvrière.

Il s'agissait d'établir un test de niveau de connaissances acquises et la justesse du rappel de ces connaissances.

Les premiers travaux ont donc été entrepris en suivant le programme de technologie générale et celui de technologie de construction, appliqués dans un groupe d'écoles de formation professionnelle.

Cinquante questions ont d'abord été retenues et groupées par famille : métallurgie, mesure-vérification, machines et outillage, technologie de construction et électricité générale. Après les premières études expérimentales, quarante questions ont été définitivement mises au point et classées en ordre de difficulté croissante.

La première expérience du test a été effectuée dans une école de formation professionnelle, pendant le dernier trimestre de scolarité. Les résultats suivants ont été obtenus, montrant nettement la progression des différents niveaux d'acquisition :

	Nombre de cas	*Moyenne de bonnes réponses*	*Ecart étalon*	*Temps moyen*
1re année	36	13,36	3,80	27' 23''
2e année	40	24	3,19	31' 10''
3e année	45	26,95	3,73	29' 44''
4e année	12	32,73	4,12	27' 19''

Bibliographie :

Bulletin C.E.R.P., n° 4, oct., déc. 1959.
G.-M. Alquier, J. Demiet et Y. Gateau : Etude d'un test de discrimination de connaissances techniques (profession de la mécanique) D.C.T.A.

ANNEXE 14 **Indicateurs d'apprentissage pour les activités de communication**

Nom : _____
Prénom : _____
Degré : _____

Activités de communication
Fiche individuelle d'appréciation
pour le cycle 3P-4P

Projet réalisé par J. Weiss, IRDP, Neuchâtel, Suisse

Indicateurs d'attitude et d'apprentissage	Appréciations	Niveau d'exigence						
		1° trim 3P	2° trim 3P	3° trim 3P		1° trim 4P	2° trim 4P	3° trim 4P
Expression écrite								
Sait produire des effets amusants avec des mots et des rimes	sans difficulté							
	avec quelques difficultés							
	avec grandes difficultés							
Ecrit des textes narratifs en respectant la chronologie des faits	sans difficulté							
	avec quelques difficultés							
	avec grandes difficultés							
Répond à un ami ou un camarade (correspondant) en tenant compte des informations ou sentiments exprimés par lui dans une correspondance précédente	la plupart du temps							
	assez souvent							
	quelques fois seulement							
Sait décrire avec précision un lieu, un objet ou un personnage	sans difficulté							
	avec quelques difficultés							
	avec grandes difficultés							
Peut rédiger sans ambiguïté les étapes d'une activité déterminée (règle d'un jeu, recettes, mode d'emploi d'un appareil)	sans difficulté							
	avec quelques difficultés							
	avec grandes difficultés							
Sait écrire un texte sous dictée	presque sans erreurs							
	avec quelques erreurs							
	avec de nombreuses erreurs encore							

Planche n° 7 : EVALUATION FORMATIVE

(Tiré de CARDINET)

ANNEXE 8

CONTROLER LA MAITRISE D'UN OBJECTIF TERMINAL GLOBAL OU D'INTEGRATION

1. Nature de l'épreuve

1.1. Ce type d'épreuve doit être construit pour évaluer un objectif terminal global ou d'intégration (OTI). Un OTI se constitue comme suit :

1) Il s'agit d'une compétence ou d'un ensemble de compétences **intégrant** (et **non juxtaposant**) les savoirs, savoir-faire et savoir-être de l'apprentissage.

2) Ces compétences s'exercent sur des situations-problèmes comprenant tant de l'**information essentielle** que de l'**information parasite** comme dans la plupart des situations de la vie ; il importe donc de pouvoir opérer au préalable un tri pertinent entre les différents types d'information, en référence au problème posé.

3) Ces compétences sont essentiellement des démarches de pensée, d'action ou d'être qui sont **significatives** pour l'apprenant (il en voit les fonctions) et qui, très vraisemblablement, laisseront des effets à long terme.

1.2. Ce type d'épreuve ne peut être donné que s'il y a eu **un apprentissage progressif et intensif à l'intégration**. A la fin de l'apprentissage, l'apprenant a travaillé régulièrement sur ce type de situations intégrées.

1.3. Etant donné le caractère situationnel des objectifs terminaux d'intégration, on pourra facilement recourir à une **évaluation critériée**. Il importe de distinguer soigneusement entre critères minimaux (permettant la certification de la réussite ou de l'échec) et critères de perfectionnement (permettant -si nécessaire- le rangement des élèves en fonction de la quantité et de la qualité des critères maîtrisés).

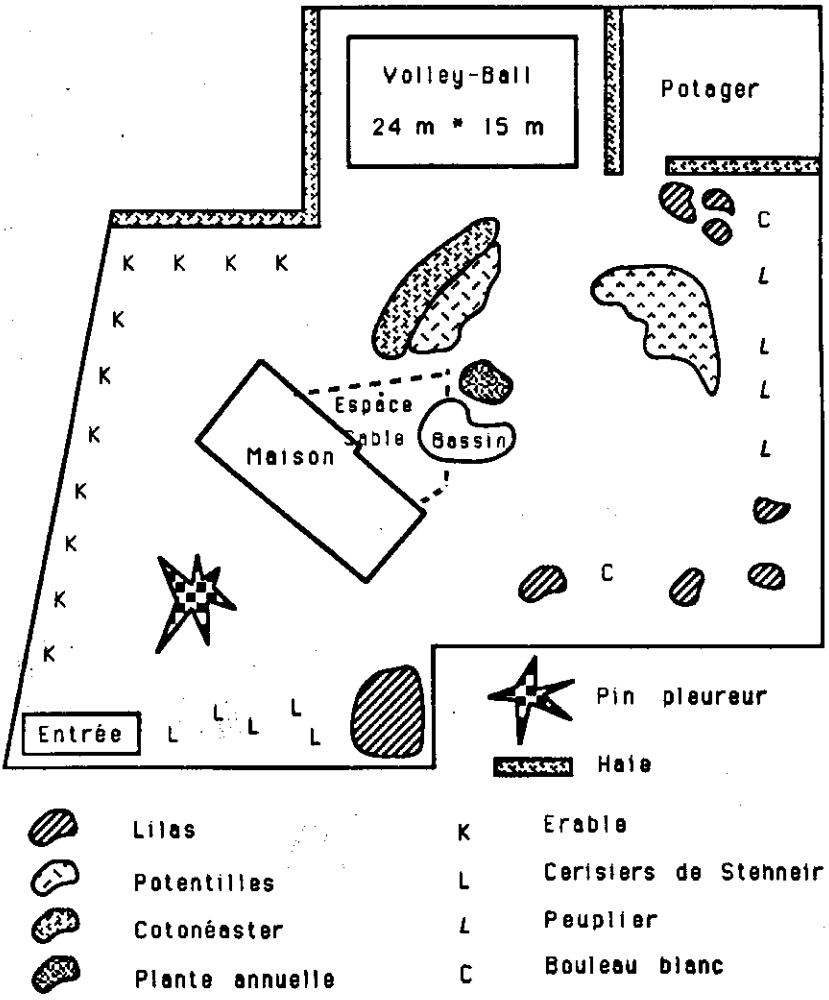
Planche n° 8 : EVALUATION DE MAITRISE PAR OBJECTIF

(Tiré de CARDINET)

1.4. Les activités proposées aux apprenants en cours d'apprentissage sont conçues dans le même esprit que les épreuves d'évaluation. L'enseignant habitue les élèves à trouver eux-mêmes les critères d'évaluation et recourt à l'auto-évaluation, une fois les critères établis. Cette procédure s'est révélée particulièrement efficace et comme le moment-clé de l'apprentissage. L'évaluation est alors de nature essentielleemnt formative.

2. Exemple d'épreuve

2.1. OTI



2.2. SITUATION

- le plan d'une propriété (cfr. ci-avant)
- une liste de prix des arbustes et de matériaux

2.3. MATERIEL dont disposent les élèves : règle millimétrée, gomme, crayon, bic, papier calque, dictionnaire, feuilles de brouillon.

2.4. QUESTION 1 : Calcule la longueur réelle totale des haies (Travaille d'abord sur une feuille de brouillon, puis
a) recopie ci-dessous, dans le cadre réservé à la réponse, la longueur totale des haies.
b) écris toutes les étapes que tu as dû suivre pour arriver à la réponse, en numérotant chacune d'elles, dans l'ordre et sans oublier).

Réponse 1 : a) La longueur totale des haies est: chiffre ..... unité .....

b) Etapes suivies

Autres questions ...

2.5. GRILLE DE CORRECTION

O.T.I.1.	*Jacqueline*	*Paul*	*Rosalie*	*Benoît*	*Christine*	*Marc*	*Véronique*
Q 1	M PP	M P	M PP		M P		M
a 1. chiffre ( )	+	-	+	-	-	-	-
2. unité correcte	+	+	+	+	+	-	+
3. unité réaliste	+	+	+	+	+	-	+
b 1. point de départ : le terrain de volley-ball	+	+		-	+		+
2. prise de mesure précise du terrain	+	+			-		+
3. coefficient de réduction	+	? -			+		+
4. prise de mesure précise (haie)	+	+			-	+	+
5. addition correcte	+	+			+	+	+
6. plan * coef. réduct.	+	+			+		+
7. multiplication correcte	+	+			+		-
8. organ. logique étapes	+	+			+		
9. lecture correcte quest.	+	+			+	-	+
10. autres procédures			+				

La première colonne reprend les critères opérationnels d'évaluation. Il faut passer beaucoup plus de temps à la préparation de la correction qu'à la correction elle-même.
Le signe + = quand le critère est observé correctement
Le signe - = quand la réponse est incorrecte
Les "blancs" = quand il n'y a pas de réponse ou quand la réponse ne permet pas de dire si le critère a été correctement observé.

Tiré de : De Ketele, J.-M. Evaluation des objectifs d'intégration,
In : *INRAP Actes de la session évaluation du 27 au 30 septembre 1983,*
Dijon: Institut National de Recherches et d'Applications Pédagogiques, 1983, Tome 1, p. 87-112

– *Chez le même éditeur* –

Collection : Orientations

animée par G. PIHOUEE – 1C –

G. CHEVALIER – Pour une clinique de groupe en psychologie de l'éducation (1C01)

A. CAROFF – L'organisation de l'orientation des jeunes en France (1C02)

D. PEMARTIN, J. LEGRES – Les projets chez les jeunes (1C03)

F. DANVERS – Le conseil en orientation en France (1C04)

J. GUICHARD – Orientation Educative de la 6ème à la 3ème (1C05)

P. VRIGNAUD – Procédures de bilan et projet formation (1C06)

M. HUTEAU – Informatique en orientation (1C07)

A. KOKOSOWSKI – Analyse et typologie des étudiants (1C90)

J.B. DUPONT – Le lycéen face à l'enseignement supérieur (1C91)

J. PELLERANO – Insertion sociale et prof. des jeunes (1C98)

Collection : Psychologie et Pédagogie du Travail

animée par B. GILLET – 1B –

B. GILLET – Comprendre les difficultés à bien résoudre les problème (1B01)

C. DIEZ – Méthode et techniques en psychologie (1B02)

B. GILLET – L'homme et son travail: traité de psychologie ergonomique (1B03)

R. FRANCES, J.L. MOGENET – Motivation et satisfaction au travail (1B04)

T. HUYNEN – Le leadership motivant pour une participation efficace (1B05)

C. DIEZ – Les éléments de psychologie générale (1B06)

Collection : Psychologie et Pédagogie du Sport

animée par A. VOM HOFE – 1F –

A. VOM HOFE, R. SIMONNET – Recherches en psychologie du sport (1F01)

A. VOM HOFE – Traitement de l'information dans les activités sportives (1F02)

L'évaluation est en crise. Tout le monde le sait, tout le monde le dit, tout le monde l'écrit, tout le monde le déplore. Bien sûr il n'existe pas de solution miraculeuse à ce problème central de l'éducation initiale et de la formation continue. Ce n'est pas une raison suffisante pour rester inactif.

Dans une première partie, l'auteur tente de dessiner les contours de ce qu'est l'évaluation, analyse ses fonctions et découvre ses mécanismes. Les différents plans de critique de l'évaluation sont passés en revue et développés. L'alternative encore utopique de l'évaluation formative est décrite en détail. Une première manière d'initialiser un changement positif, une amélioration technique significative, un décentrage par rapport aux routines est proposée: l'utilisation raisonnée et raisonnable des tests pédagométriques.

L'objectif du livre est donc de proposer une réflexion théorique qui ressaisisse les fondements de la problématique évaluative et de prolonger cette réflexion au niveau de la pratique par l'usage modéré et justifié des techniques de l'évaluation scientifique.

La combinaison de la réflexion et de l'information concrète facilitera chez les praticiens de la pédagogie et de l'évaluation scolaire et professionnelle qui en éprouvent le besoin la mise en route d'un processus dynamique de changement, de différenciation, d'élargissement, d'enrichissement et d'amélioration des pratiques évaluatives.

Après avoir travaillé à l'Education Nationale comme instituteur puis comme psychologue, Daniel PASQUIER est chargé d'études à l'Association nationale pour la Formation Professionnelle des Adultes. Chercheur en psychopédagogie il s'intéresse actuellement aux grandes questions de l'éducabilité des adultes en formation et au travail: mesure du potentiel d'apprentissage, outils et démarches de remédiation cognitive et socio-affective, étude des effets psychologiques et sociaux des interventions de remédiation...