

UNIVERSITÉ DE FRIBOURG

Institut de Pédagogie

**LES INNOVATIONS SCOLAIRES ROMANDES
ET LEUR ÉVALUATION**

**Cours donné en 1993-1994
par Jean Cardinet
avec la participation de Raymond Clavel
et des exposés pris en charge par des étudiants**

DESCRIPTION DES COURS

1. CARDINET Jean

2. **Les innovations scolaires romandes et leur évaluation**

3. Mercredi 17 h.-19 h.

4. Début: 20.10.93

5. ~~222~~ Salle C-0.106

6. Si l'on veut évaluer le fonctionnement d'une institution, ou la réussite d'une innovation, il faut expliciter le cadre de pensée dans lequel on se situe, puis le modèle d'évaluation que l'on choisit. Ces choix dépendent du référentiel par rapport auquel on veut apprécier les résultats.

Les recherches qui ont accompagné l'introduction des innovations scolaires romandes serviront à illustrer les démarches évaluatives possibles et à faire apparaître les problèmes qu'elles posent.

L'expérimentation de la méthode de lecture, du nouveau curriculum de mathématique et du nouvel enseignement du français, mettront en lumière trois modèles d'évaluation très différents, applicables à des curriculums entiers. Des innovations de moindre envergure demandent des démarches évaluatives plus limitées: des exemples en seront donnés également.

Le cours s'adresse à l'ensemble des étudiants en pédagogie. Les étudiant(e)s du 1er Cycle, ou en branche secondaire, seront évalués en fin d'année sur leur connaissance des principes pédagogiques des innovations présentées et leur compréhension des référentiels des méthodes utilisées.

Les étudiants du 2e Cycle en branche principale travailleront le même sujet au niveau taxonomique de l'analyse: ils présenteront des exposés dans le cadre du cours, pour expliquer les objectifs des innovations effectuées, résumer les résultats obtenus, et discuter les démarches de recherche utilisées.

1ÈRE PARTIE: THÉORIE DE L'ÉVALUATION DE CURRICULUM

CH. 1. LE CONCEPT DE CURRICULUM: PRÉSENTATION THÉORIQUE

1. Origine

Le terme de "curriculum" apparaît au 17e siècle dans le contexte des efforts de réforme des universités, suscités eux-mêmes par le mouvement de la Réforme. Il correspond donc à une structure organisationnelle imposée par les autorités, dans le but de mettre de l'ordre dans la conduite des études.

Progressivement, cependant, le sens du terme s'est élargi. Si on n'englobait au départ sous ce terme que la succession des cours, on y a ajouté d'abord toutes les expériences éducatives (par ex. non-verbales, pratiques), introduites volontairement par les enseignants.

A ce curriculum explicite est venu ensuite s'ajouter le curriculum caché, qui correspond à tous les apprentissages non planifiés, parfois même non souhaités, que l'école suscite.

Le curriculum explicite peut être traité à deux niveaux: idéal (c'est le curriculum prévu) et réel (le curriculum réalisé); {la différence entre les deux tient au degré d'implémentation du curriculum prévu, qui varie en fonction de l'attitude des enseignants et des autres conditions qui affectent l'enseignement}. Le degré de réceptivité des élèves peut encore varier, ce qui détermine alors le curriculum reçu.

On voit donc que le terme est maintenant appliqué à des réalités très variées. On ne s'étonnera pas de le voir désigner tantôt l'ensemble de toutes les expériences éducatives, tantôt un programme d'apprentissage pour une branche particulière.

2. Définition

Le curriculum, dans son acception la plus globale, est le plan, ou le programme, de toutes les expériences que l'école fait vivre à l'apprenant.

Karl Frey en donne une définition un peu plus détaillée: "Le curriculum, c'est une conception de l'enseignement qui en fait, à l'intérieur d'une période donnée, un système cohérent, comportant des procédures multiples pour la planification, la réalisation optimale et le contrôle des résultats."

3. Eléments constitutifs

- 1) contexte: les expériences des apprenants s'insèrent nécessairement dans un cadre historique, géographique, social, psychologique, etc. qui modifie les effets de ces expériences.
- 2) finalités du curriculum global: l'apprentissage d'une branche particulière n'est pas une fin en soi pour un éducateur; il vise à aider l'enfant, ou l'apprenant, à se développer dans une certaine direction.

- 3) objectifs de la discipline: la nature des expériences prévues est liée à une certaine conception de la discipline et de son apprentissage, qui s'exprime dans les objectifs terminaux visés.
- 4) progression didactique: les savoirs et savoir-faire enseignés sont organisés selon une séquence particulière, considérée comme la plus efficace.
- 5) moyens d'enseignement qui matérialisent et soutiennent cette progression didactique: manuels, cahiers d'exercices, fiches de travail, méthodologie, etc.
- 6) transactions, échanges de natures diverses entre l'enseignant et sa classe.
- 7) moyens d'apprécier les résultats de ces expériences, attendus et parfois inattendus.
- 8) moyens de remédiation.

4. Exemple: les méthodes de lecture américaines

Lire: Cardinet, J. Des exemples de curriculums: les méthodes de lecture américaines. *Educateur*, N°39 (1972), p. 931-934.

5. Bases scientifiques

Une progression aussi bien étudiée doit s'appuyer sur une théorie scientifique de l'apprentissage. On peut en considérer quatre:

5.1. La perspective behavioriste

L'acquisition du savoir est influencée par le contexte, en particulier par les renforcements qui suivent nos actions: nous apprenons à ne plus toucher ce qui nous a brûlé, par exemple. Enseigner efficacement revient alors à renforcer chaque réponse qui va dans la bonne direction, en vérifiant chaque petite étape aussitôt que possible. La progression doit être graduelle pour que l'élève ne soit jamais arrêté, mais réussisse au contraire constamment. Cette réussite doit maintenir sa motivation. En élargissant progressivement le répertoire des problèmes qu'il pose, l'enseignant arrive à faire traiter des cas de plus en plus complexes, sans que l'élève ne fasse d'erreurs en chemin. L'apprentissage est donc conçu comme linéaire.

Les curriculums qui s'inspirent de cette perspective contiennent des leçons centrées sur un savoir faire défini, construit de façon progressive, selon un cheminement systématique, avec des tests formatifs très fréquents, permettant de revoir et d'automatiser les connaissances. L'intérêt particulier de cette pédagogie est de permettre à chaque élève d'aller à son allure propre, mais de toujours réussir finalement, à condition de répéter les séquences d'apprentissage un nombre de fois suffisant.

En fait, ces progressions ne conduisent pas toujours à une capacité de généralisation suffisante: fréquemment, l'élève ne transfère pas bien ce qu'il a appris dans un cours programmé à des situations nouvelles, parce qu'il n'a pas travaillé à un niveau d'abstraction suffisant et n'a pas construit des concepts assez généraux. D'autre

part, ce pilotage du comportement par l'extérieur peut conduire à des tentatives de manipulation que des éducateurs ne peuvent accepter.

5.2. *La perspective développementaliste*

Pour cette école de pensée, développée d'abord par Piaget, c'est dans l'individu lui-même et non à l'extérieur que l'on peut situer les déterminants de l'apprentissage. Les enfants organisent leur représentation du monde sur la base de leur activité et construisent des synthèses provisoires (les stades logiques) relativement cohérentes, qui les rendent capables de maîtriser leur environnement. Ces synthèses les rendent peu sensibles, par contre, à des éléments qui contredisent leurs conceptions (ce qui justifie d'adapter les programmes au niveau de développement cognitif des élèves).

Les curriculums qui s'appuient sur une conception développementaliste y trouvent un fondement théorique pour déterminer l'âge optimum de présentation des notions théoriques et le type d'activité (et notamment de manipulation) qui permet de les assimiler. Le rôle de l'enseignant devient aussi différent: ce dernier est moins le pourvoyeur d'informations que l'animateur de la réflexion des enfants, celui qui encourage les efforts d'exploration, l'interaction entre pairs et la réflexion personnelle.

5.3. *La perspective cognitiviste structuraliste*

Cette tradition reconnaît aussi bien le rôle de la répétition des associations et du renforcement dans le développement de l'apprentissage, que celui de l'activité mentale qui aboutit à la construction de concepts et de cartes mentales. Poursuivant les travaux de l'école "Gestaltiste", cette perspective souligne l'importance de la structure des informations présentées: on perçoit d'abord les structures dans le monde qui nous entoure et on mémorise ces structures. Les situations d'apprentissages doivent donc être préstructurées par l'enseignant pour conduire à des découvertes chez l'enfant. L'objet de la didactique doit être l'organisation cohérente de ces structures successives, de manière à permettre des "insights" donnant accès à de nouveaux concepts.

En conséquence, apprendre une discipline, c'est apprendre comment des concepts sont reliés, dégager les structures qu'ils forment. Le langage peut souligner ces structures, par exemple lorsqu'on propose aux élèves des "organiseurs avancés" qui annoncent les idées principales avant qu'on les développe. D'autres formes de présentation (visuelles, auditives, etc.) peuvent jouer sur la façon dont chaque élève structure le mieux son environnement.

Les implications pour la construction des curriculums sont nombreuses, parce que cette approche permet d'intégrer plusieurs des propositions des deux approches précédentes, dans son programme d'apprentissage par la découverte: il s'agit à la fois de manipuler l'environnement et de stimuler l'activité cognitive personnelle. On peut rattacher à cette approche les jeux de

simulation, les activités de recherche, les travaux pratiques dirigés, la construction de cartes cognitives, la prise en compte des contextes culturels d'origine des élèves, etc.

5.4. *La perspective cognitiviste par traitement de l'information*

Cette approche s'accorde avec la précédente sur la majorité de ses conclusions, mais les justifie souvent différemment.

Pour ses théoriciens, l'enfant est équipé pour mettre de l'ordre dans la réalité en dégagant des ressemblances. Il en tire des cadres de référence, des "schémas", qui lui permettent de se retrouver et d'agir. Mais tout le contexte intervient dans ces représentations, et notamment la signification personnelle et sociale de la situation: l'enfant cherche à comprendre et à s'adapter. Il reconstruit donc ce qu'il perçoit et ce qu'il mémorise à partir de toute son expérience passée.

C'est le but des cognitivistes de le suivre dans cette reconstruction, de déterminer par quels processus il y parvient. L'étude de l'apprentissage d'une tâche ne peut être fructueuse tant que l'on ne sait pas comment le sujet réalise cette tâche. C'est pourquoi la comparaison des démarches des novices et des experts est une des voies d'approche favorites de cette école.

Pour guider la construction des curriculums, ces études montrent la nécessité d'automatiser certains traitements d'information de niveau bas (les opérations arithmétiques, par ex.) pour laisser l'esprit libre pour des tâches cognitives plus complexes, de résolution de problèmes notamment. Mais la pensée a besoin de s'appuyer sur des connaissances spécifiques pour raisonner, ce qui justifie l'apprentissage de disciplines structurées et de réseaux de connaissances. Il semble même qu'il existe autant d'intelligences que de cadres d'action et de pensée, ce qui conduit à diversifier les domaines d'activités abordés durant la scolarité.

6. Bases humanistes

6.1. *Les risques de la rationalisation*

Il faut veiller, cependant, à ne pas tomber dans un excès de scientisme. Certains créateurs de curriculums envisageaient l'école comme une usine dont ils voulaient augmenter la productivité. Le risque de cette métaphore est de déshumaniser l'éducation, d'une part en ne prenant en compte que l'aspect rationnel de l'éducation, d'autre part en donnant plus d'importance aux moyens qu'aux fins.

6.2. *Critiques à l'encontre des curriculums courants*

Les considérations économiques l'emportent souvent sur les considérations esthétiques, éthiques et spirituelles. Les besoins des milieux socioéconomiques défavorisés ne sont pas pris en compte, ni ceux des minorités culturelles (enfants migrants, par ex.). L'idéologie de la classe dominante est implicitement acceptée et propagée. De plus, les enseignants sont privés de leur rôle, transformés en ouvriers spécialisés, "prolétarisés".

6.3. Nécessité de plus d'originalité

La connaissance, même scientifique, est une construction sociale. Il est nécessaire de repenser l'éducation tout autant que la société, (en évitant les préjugés sexistes, ou nationalistes, ou racistes, par exemple). Il est important de permettre aux enfants de s'exprimer et de se développer, sans leur imposer un moule préétabli. Tout ce qui se passe en classe n'a pas besoin d'être prévu dans le curriculum...

6.4. Quatre types de curriculums

Il existe toujours, en pédagogie, une tension entre ceux qui veulent garder le curriculum centré sur l'apprentissage des disciplines traditionnelles (les traditionalistes) et ceux qui veulent le changer. Parmi ces derniers, on peut distinguer ceux qui veulent mieux prendre en compte les besoins psychologiques de l'individu, (les développementalistes, selon les conceptions de Dewey, vers 1900), et ceux qui sont motivés par des justifications d'ordre sociologique (les réformistes, comme Bobbitt aux USA vers 1920). Mais ces trois écoles pédagogiques s'opposent à la tendance des techniciens qui laissent le choix des options philosophiques aux politiciens, et qui se satisfont de poursuivre l'efficacité, en veillant seulement à ce que le curriculum atteigne les buts qui lui sont assignés. (Edward Thorndike ou Robert Glaser). (Il ne s'agit, bien sûr, que de priorités relatives, car chaque curriculum doit bien tenir compte de tous les aspects éducatifs: culturels, psychologiques, sociologiques, et didactiques.)

7. Principes de base pour la construction d'un curriculum

Ralph Tyler publia en 1949 "Basic principles of curriculum and instruction", dont la démarche s'inspire de quatre questions:

- Quels sont les buts éducatifs que l'école cherche à atteindre?
- Quelles expériences éducatives peut-elle offrir qui ont des chances de conduire à ce résultat?
- Comment ces expériences éducatives peuvent-elles être organisées pour être efficaces?
- Comment pouvons-nous nous assurer que les buts visés ont été atteints?

Il s'agit d'une démarche de bon sens, qui s'appuie surtout, en fait, sur l'analyse des dysfonctionnements des curriculums en usage, pour proposer des améliorations.

Ce n'est pas faire preuve d'étroitesse d'esprit que d'accepter cet effort de rationalisation.

Bibliographie

Darling-Hammond, L., & Snyder, J. (1992). Curriculum studies and the traditions of inquiry: the scientific tradition. In P. W. Jackson (Eds.), Handbook of research on curriculum (pp. 41-78). New York: Macmillan.

Jackson, P. W. (1992). Conceptions of curriculum and curriculum specialists. In P. W. Jackson (Eds.), Handbook of research on curriculum (pp. 3-40). New York: Macmillan.

Des exemples de curriculums : les méthodes de lecture américaines

Les travaux de la CIRCE et les efforts de coordination romande en général ont mis la réforme des programmes à l'ordre du jour. En prenant connaissance des textes, des articles ou des exposés consacrés à ce sujet, chacun a pu rencontrer des termes nouveaux, comme idée directrice, objectifs opérationnalisés, séquence d'apprentissage, instruments d'évaluation, mesures correctrices, et plus généralement ce mot clé qui semble les englober tous : celui de *curriculum*. De quoi s'agit-il ?

Les partisans du latin peuvent marquer un point : le recours aux sources de notre culture permet aux Suédois, aux Américains, aux Allemands ou aux Italiens d'utiliser le même mot et de s'entendre sur le même concept : « Le curriculum, est une conception de l'enseignement qui en fait, à l'intérieur d'une période donnée, un système cohérent, comportant des procédures multiples pour la planification, la réalisation optimale et le contrôle des résultats. » (Karl Frey, Kriterien in der Curriculumtheorie und -strategie in Kriterien in der Curriculumkonstruktion, Weinheim - Basel, 1970, Beltz.)

La caractéristique principale d'un curriculum par opposition aux plans d'études anciens, c'est d'être très cohérent et systématique et d'organiser la situation d'apprentissage de l'élève en ne laissant rien au hasard. On précise d'abord les comportements attendus de l'élève en fin d'étude, en veillant à assurer la cohérence entre les objectifs généraux de l'enseignement et les objectifs particuliers de la discipline enseignée. On recherche ensuite la progression la meilleure en fonction de la population d'élèves envisagée pour effectuer cet apprentissage : on définit une séquence d'enseignement et son déroulement dans le temps. Pour chaque étape, on décide des stratégies que l'on suivra pour guider rationnellement le travail des élèves ; on détermine les méthodes et les moyens d'enseignement les plus appropriés ; on construit les matériels éducatifs nécessaires. Ceci amène généralement à prévoir aussi bien les procédures d'individualisation que les activités de groupe qui assureront l'assimilation et maintiendront la motivation. Des épreuves de contrôle doivent alors permettre de vérifier à chaque étape les progrès de l'élève et de déceler ses difficultés. Enfin, des procédures d'appui et de rattrapage viennent suivre ces contrôles pour assurer les corrections (individuelles et collectives) qui s'avèrent toujours nécessaires. Au total, le curriculum doit permettre d'atteindre

les objectifs de départ : une série d'expérimentations préalables donnent la possibilité de l'adapter jusqu'à ce que son efficacité soit garantie.

La mission confiée à l'IRDP d'expérimenter la nouvelle méthode de lecture « S'exprimer - Lire » nous a obligé à recueillir une large documentation préalable sur ce sujet. C'est ainsi que nous avons pris connaissance des moyens utilisés à l'étranger pour cet enseignement. Les méthodes américaines que nous avons examinées (celles de Ginn et de Hoffmann en particulier) étaient conçues comme des curriculums et nous ont paru représenter d'intéressants exemples concrets de ce que pouvait produire une approche systématique de la pédagogie. C'est à ce titre que nous allons en présenter une description générale, non pas pour affirmer que ces méthodes sont bonnes en elles-mêmes, et encore moins pour proposer de les transposer en français. Nous souhaitons simplement apporter une documentation sur la construction des curriculums, qui soulèvera sans doute des prises de position divergentes, mais qui, par là même, nous paraît un point de départ intéressant de réflexion pédagogique.

La simple définition abstraite d'un curriculum, telle que nous l'avons esquissée ci-dessus intéresse ou inquiète peu, parce qu'elle semble irréalisable : les exemples vécus sont d'un autre pouvoir...

Les objectifs

La première étape de tout curriculum est de préciser le résultat qu'il cherche à atteindre. Ceci implique, bien sûr, comme dans nos plans d'études, une certaine analyse de la matière à enseigner, une certaine vue théorique du domaine, telle que le spécialiste des mathématiques, de la linguistique, de la géographie peut se le représenter ; mais on n'en reste pas à ce catalogue de notions théoriques. Le résultat désiré est précisé en termes de comportements de l'élève. Voyons ce que ceci signifie pour la lecture.

On commence par définir ce qu'est la lecture. Savoir lire c'est :

1. percevoir clairement les symboles ;
2. comprendre les mots, les phrases et les textes ;
3. interpréter correctement leur signification ;
4. employer couramment les sources d'information écrite.

De cette définition très large découle-

ront toutes les décisions ultérieures relatives à la méthode, et en premier lieu le choix des objectifs comportementaux recherchés qui se classent sous six têtes de chapitres :

1. possession des prérequis ;
2. capacité de déchiffrer les mots ;
3. capacité de comprendre un texte ;
4. capacité de lire de façon créative ;
5. intérêt pour la lecture ;
6. maîtrise du langage.

Ces objectifs sont ensuite précisés année par année, le curriculum couvrant l'enseignement de la lecture à travers l'ensemble de la scolarité pré-obligatoire et obligatoire. Par exemple pour la première année de l'école obligatoire on précisera ainsi la capacité de déchiffrer les mots : c'est deviner leur sens, les reconnaître globalement, les analyser phonétiquement, analyser certains aspects de leur structure, et les décomposer en syllabes. Pour chacune de ces capacités, une analyse plus poussée est encore réalisée à l'intérieur du livre du maître, et précisée semaine par semaine sous forme de tableaux synoptiques à double entrée, les objectifs (en colonnes) étant mis ainsi en correspondance avec les activités des élèves (en lignes). Pour chaque leçon enfin, l'enseignant reçoit une indication précise des objectifs visés par chaque élément du travail proposé.

La stratégie

L'effort de cohérence dans la conception et dans l'exécution qui caractérisent la mise au point des curriculums serait peine perdue si les enseignants n'étaient pas informés dans le détail de cette stratégie. Ils risqueraient facilement de dénaturer la progression prévue en attribuant une signification différente aux exercices proposés ou en adoptant une attitude en contradiction avec les objectifs désirés.

C'est pourquoi le livre du maître est très détaillé. Il comporte 430 pages par exemple pour le premier semestre d'apprentissage de la lecture, qui comporte 45 leçons. Le premier chapitre (11 pages) présente le curriculum et ses caractéristiques particulières ; le deuxième présente la conception que se font les auteurs de la lecture et de l'apprentissage de la lecture (26 pages) ; le troisième présente les objectifs de l'année et la méthodologie correspondante (27 pages). Le reste du livre comporte des indications ou suggestions relatives à chaque leçon (en moyenne 7 pages par leçon).

Il n'est pas possible de rendre compte dans le détail du contenu de cette information préalable destinée aux enseignants. Notons-en seulement l'organisation et les

idées principales qui montreront à quel niveau se situe d'emblée cette préparation théorique générale.

1. Fondements psychologiques

La lecture est un acte complexe, qui exige beaucoup plus que de faire correspondre des sons à des lettres. Il faut comprendre les idées et même les dépasser pour en voir les implications. Il faut utiliser la lecture pour étudier, se distraire, etc.

La lecture est un acte de communication. Développer la capacité d'échange sur le plan oral constitue donc un soutien nécessaire à l'apprentissage de la lecture, car une communication n'est motivée que si elle est possible dans les deux sens.

L'apprentissage de la lecture ne se fonde pas sur un mécanisme unique ; c'est pourquoi il faut combiner plusieurs approches, globales et synthétiques.

L'apprentissage doit être significatif pour l'enfant. Le contenu du manuel doit donc être relié à ses intérêts, à son expérience, à ses préoccupations.

Des activités motivées conduisent à des apprentissages simultanés, sur plusieurs plans à la fois (décodage, vocabulaire, pensée logique, intérêts, apprentissage social, par exemple). Il faut donc créer une atmosphère plaisante stimulante mais non contraignante, où l'élève agisse librement.

Le rythme et la façon d'apprendre diffèrent d'élève à élève : il faut prévoir les moyens pour s'adapter à ces différences.

Les prérequis existent à tous les niveaux de l'apprentissage. Tout apprentissage complexe repose sur des apprentissages plus simples. Il faut donc toujours assurer les bases, mais également préparer la suite. C'est pourquoi la continuité est nécessaire au curriculum. L'activité de réflexion sur les textes proposés, par exemple, qui peut surprendre en première année, vise à préparer les élèves à une lecture intelligente par la suite.

L'apprentissage de la lecture est un processus continu. Il faut prévoir, pour chaque capacité à développer, son introduction et sa consolidation, selon une progression méthodique, toutes les parties du curriculum se répondant et se soutenant les unes les autres.

2. Fondements pédagogiques

Les activités de lecture doivent être riches et variées pour atteindre les objectifs multiples posés au départ. Le curriculum de Ginn propose quatre programmes parallèles. Le premier vise à développer les aspects intellectuels de la lecture, le deuxième la motivation à la lecture, par des activités de loisir, le troisième l'habitude d'utiliser la lecture, par

un choix d'activités parascolaires et le quatrième est un programme éducatif où la lecture sert à développer la connaissance de soi-même, des autres, du monde et des valeurs.

Les activités de lecture doivent être reliées aux autres domaines de l'enseignement. C'est pourquoi des lectures à thèmes scientifiques, géographiques, historiques et biologiques sont prévues, pour motiver l'introduction de leçons sur ces sujets. Des exercices de lecture de tableaux, de cartes, de graphiques développent la notion trop étroite de lecture de phrases. Les exercices d'analyse de la situation d'un texte préparent enfin l'étude des manuels de toutes les autres disciplines.

Les résultats doivent être évalués de façon continue. L'accent est mis sur une évaluation continue plutôt que périodique, prise en charge par l'élève dans toute la mesure du possible, analytique plutôt que globale et centrée sur l'atteinte des objectifs pédagogiques plutôt que sur le classement des élèves les uns par rapport aux autres.

Les résultats de l'évaluation doivent servir de base à une activité correctrice : des exercices sont donc proposés en même temps que chaque test, pour aider à compenser les déficiences qui auraient été mises en évidence.

La tactique

Les principes précédents feraient aisément l'unanimité des enseignants, mais seraient de peu d'utilité si le moyen ne leur était pas donné de les faire passer dans la pratique. C'est à ce niveau de la réalisation concrète que l'on attend, sans doute avec quelque scepticisme, celui qui veut réaliser un curriculum. Nous présenterons cette mise en œuvre d'une méthode de lecture en deux parties, en indiquant d'abord le contenu de la leçon type, puis en situant cette leçon dans l'ensemble de l'activité pédagogique.

1. L'ACTIVITÉ EN CLASSE SELON LA LEÇON TYPE

a) Développement de la préparation des élèves

Une première phase de la leçon vise à présenter aux élèves le cadre et le vocabulaire du texte à lire, pour assurer leur compréhension et stimuler leur intérêt. Dans ce but, le livre du maître contient un résumé des thèmes abordés et des valeurs morales implicites dans chaque leçon, qui permet à l'enseignant d'indiquer l'idée générale du passage. Tous les mots nouveaux sont annoncés au maître. On lui propose une série de ques-

tions à l'intention des élèves et qui doivent aider ces derniers à préciser eux-mêmes le sens du vocabulaire qu'ils rencontrent pour la première fois.

b) Lecture de l'histoire

Tous les textes, dès le tout premier, sont des histoires possédant un certain intérêt dramatique ou humoristique. Au départ, l'essentiel de la signification est apporté par l'image, le texte se réduisant à quelques mots associés de multiples façons et que l'enfant apprend à reconnaître globalement. La lecture est ainsi toujours porteuse de signification et jamais un exercice abstrait.

C'est ce qui permet au maître de faire lire le texte individuellement et silencieusement, dès le début de la méthode. Cette lecture est stimulée de plus par un certain nombre de questions (formulées dans le livre du maître). Les enfants cherchent dans l'image, et plus tard dans le texte, la réponse à ces questions liminaires. La lecture est ainsi toujours motivée, apporte une réponse à une attente, l'intention pédagogique étant évidemment de donner l'habitude d'une lecture attentive, active et critique.

Sur la base de cette première lecture, les élèves répondent aux questions qui leur étaient posées, puis le texte est repris oralement par les élèves. Il ne s'agit pas cependant d'un « exercice de lecture ». Cette activité est motivée par une visée plus générale : lire l'histoire à un groupe qui ne l'a pas lue, à un visiteur, à un camarade qui doit la mimer ou l'exprimer par des marionnettes, citer les parties que l'on trouve les plus intéressantes, celles qui expriment telle idée, celles qui vérifient telle interprétation, contiennent tel mot, etc. La relecture est liée donc à l'analyse et à l'interprétation.

c) Exercice des capacités nécessaires à une bonne lecture

Une série d'exercices sont offerts ensuite à la maîtresse pour faire acquérir les connaissances techniques indispensables. Une partie d'entre eux sont déjà préparés dans des cahiers d'exercices, établis de façon à correspondre aux thèmes et au vocabulaire de chaque leçon, mais conçus en même temps pour atteindre les objectifs particuliers de cette leçon, selon la séquence générale d'apprentissage dont nous avons déjà parlé. D'autres exercices sont proposés dans le livre du maître, de façon à laisser à l'enseignant la possibilité d'une adaptation personnelle et à assurer une certaine variété dans son travail. Des exercices de vocabulaire et des exercices de compréhension viennent enrichir la connaissance de la langue écrite.

d) Expérience linguistique

Le texte sert ensuite de point de départ pour des activités linguistiques beaucoup plus globales : analyse du texte de divers points de vue, choix après discussion en groupe d'activités parascolaires liées au thème de la lecture, expression d'opinions individuelles sur le sujet traité, présentation d'expériences individuelles en rapport avec ce thème, invention d'histoires analogues, expression dramatique et plus tard expression écrite de ces histoires.

e) Activités d'enrichissement

On évite autant que possible de proposer aux élèves des activités dispersées, peu motivantes dans la mesure où elles paraissent artificielles. Au contraire, on encourage l'intégration de toutes les activités de la classe autour d'un thème, d'un but commun. Pour chaque unité d'enseignement, correspondant à 5 à 8 leçons, ou à une période d'une à deux semaines, une activité de coordination est proposée : ce peut être d'écrire un journal de classe, de préparer une exposition, de préparer une fête, etc. D'autres idées sont offertes pour établir des liaisons entre la lecture et le bricolage, les activités artistiques, la musique, des disques, des films, même des livres d'art ou des reproductions de tableaux. Une bibliographie d'histoires à lire aux enfants ou de livres adaptés à leur niveau de lecture est également offerte aux enseignants pour enrichir leurs leçons.

2. MISE EN ŒUVRE GÉNÉRALE DE LA MÉTHODOLOGIE PROPOSÉE

La leçon type s'insère elle-même dans un cycle d'une à deux semaines : l'unité d'enseignement, et l'ensemble des unités suit un cycle annuel.

Au début d'une unité d'enseignement, le maître reçoit des indications sur les activités prévues dans l'unité et les objectifs qu'elles visent. Pour chaque leçon, il reçoit des indications supplémentaires qui lui permettent de se préparer convenablement. Des suggestions lui sont offertes de travaux en groupes, qui lui donnent la possibilité matérielle de reprendre l'enseignement individuellement avec certains élèves en difficulté après la leçon collective.

A la fin d'une unité, il est prévu de faire le point à la fois avec les élèves et avec leurs parents. Des tests de déchiffrement, de compréhension et de vocabulaire sont inclus dans la méthode. Des questionnaires conduisent la maîtresse à évaluer les attitudes, les capacités, les points forts et faibles de chaque élève. Des suggestions accompagnent ces épreu-

ves pour des activités individualisées, adaptées aux élèves soit en retard, soit en avance par rapport au niveau moyen de la classe. Des exercices de rattrapage correspondent aux déficiences éventuelles.

L'ensemble de l'année se situe entre deux prises d'information plus générales, la première contrôlant les prérequis de l'année (c'est-à-dire les aptitudes perceptives, verbales et intellectuelles sur lesquelles devront s'appuyer les apprentissages envisagés) et la seconde contrôlant les acquisitions de l'année suivante. On espère faciliter par ces tests le placement convenable des élèves dans leur niveau de lecture.

Les moyens

C'est au niveau des moyens d'enseignement fournis à l'enseignant que la compétition est la plus vive entre les diverses méthodes à disposition sur le marché. C'est en effet sur les aspects matériels que l'on s'arrête au premier abord, parce que ce sont les plus visibles et les plus faciles à contrôler. La cohérence du curriculum n'apparaît qu'à ceux qui examinent l'ensemble de la méthode beaucoup plus en détail. Sans attribuer donc une importance excessive à la richesse des matériels et des media proposés, citons ceux qui sont offerts actuellement.

1. Les manuels

Les curriculums très structurés s'appuient généralement sur des manuels de lecture qui permettent de contrôler avec précision la progression de l'apprentissage. Les livres forment une suite ininterrompue depuis le jardin d'enfants deux ans avant l'entrée à l'école obligatoire, jusqu'en cinquième année ou même plus. Ils donnent lieu à des études poussées concernant leur reliure, leur mise en pages, la taille et la forme des caractères utilisés, le style des illustrations et leur fonction par rapport au texte et aux exercices, la longueur des histoires, le vocabulaire utilisé et tous les aspects objectivement mesurables.

Le choix des thèmes, bien que plus qualitatif, a également été particulièrement contrôlé ces dernières années. On veille à l'équilibre du passé et du présent, du réel et de l'imaginaire, de la civilisation américaine et des autres, du contexte urbain et rural. Les valeurs morales visées sont explicitées, les problèmes sociaux, raciaux et économiques sont abordés avec une attention spéciale.

La motivation des élèves est un autre souci primordial. On s'appuie sur des enquêtes d'intérêts pour les âges considérés. On exploite l'humour, le « sus-

pense », la vivacité du dialogue, les illustrations. Des livres parallèles permettent de choisir les lectures demandées par les élèves et d'enrichir les programmes par des lectures individuelles.

2. Les cahiers d'exercices

Parallèlement à chaque leçon sont prévus des exercices non seulement de lecture, mais aussi de prérequis, dont le rôle est de permettre l'assimilation des connaissances, ainsi que la mise en évidence des difficultés éventuelles de l'élève.

3. Les moyens audio-visuels

Ils varient des plus simples (tableaux muraux) aux plus complexes (projecteurs individuels munis de magnétophones, lisant le texte imprimé ou donnant signes et explications) en passant par les flanellographes, les jeux de mots et de lettres, etc. L'avantage des projecteurs individuels est de permettre une véritable individualisation : chaque enfant peut progresser à son rythme puisqu'il se donne les explications à lui-même par le moyen des disques et des cassettes à sa disposition.

4. Les fichiers autocorrectifs

Des exercices de compréhension ont été préparés sur des textes de difficulté croissante. L'élève peut à chaque fois contrôler lui-même les réponses qu'il a données à la fiche précédente.

5. Les tests

Différents moyens d'évaluation accompagnent les moyens d'enseignement. En début d'année des tests de prérequis sont conçus comme des épreuves prédictives du progrès attendu pour chaque élève s'il suit l'enseignement de l'année considérée. Ils doivent permettre de situer l'élève dans le groupe le meilleur pour son apprentissage. En cours d'année, des tests « formatifs » et des questionnaires remplis par l'enseignant vérifient si les objectifs particuliers de chaque unité ont été atteints ou non. Ils ne comportent pas d'étalement car une comparaison d'élève à élève n'est pas nécessaire et ils ne portent que sur des difficultés spécifiques, examinées isolément de façon à permettre un diagnostic des difficultés de l'élève. Une comparaison plus globale n'est effectuée qu'à quelques occasions où il peut être utile de situer le niveau des élèves par rapport aux résultats d'un ensemble d'enfants du même âge. Les tests sont alors de nature « sommative », c'est-à-dire qu'ils font appel à une activité de lecture complexe, sur un texte comportant l'ensemble des difficultés déjà étudiées.

Des manuels sont offerts pour les tests prédictifs et sommatifs, fournissant les explications nécessaires pour faire passer, coter et surtout évaluer les résultats à ces épreuves. Les tests formatifs, au contraire, sont inclus dans le livre du maître et les cahiers d'exercices des élèves et ne sont accompagnés que de suggestions d'exercices de rattrapage.

6. La documentation supplémentaire

Le livre du maître contient enfin d'abondantes bibliographies commentées sur la psychologie et la pédagogie de la lecture, sur les sources d'enrichissement possibles pour son enseignement, sur des livres de poésies, de chansons, de jeux adaptés à cet âge, sur les ouvrages à conseiller pour des lectures libres, etc.

La formation des enseignants

Nous avons noté déjà le caractère systématique d'un curriculum, qui prend en charge tous les éléments favorables à l'apprentissage des élèves. La formation des enseignants ne peut donc être négligée, puisqu'elle est le facteur de réussite essentiel.

Le livre du maître commence par expliquer les fondements psychologiques et pédagogiques de la méthode et ses objectifs généraux. Il présente ensuite le niveau de départ prévu, c'est-à-dire les prérequis de l'année en question et comment les contrôler. Les objectifs annuels sont alors décrits, puis la progression en cours d'année est détaillée au moyen de tableaux synoptiques. Un chapitre explique les auxiliaires pédagogiques à disposition. Enfin la méthodologie est donnée par un plan de leçon type et des conseils généraux.

Après cette formation de départ, le livre du maître prend en charge également la préparation des leçons en cours d'année. Il fournit des indications détaillées pour chaque unité d'enseignement et pour chaque leçon, comme nous l'avons vu plus haut. Il donne une liste cumulée de tous les mots employés dans les manuels précédents et une liste séparée du vocabulaire de l'année, ce qui permet de situer le niveau de la classe et de composer des exercices adaptés à ses possibilités et également aux objectifs visés. Dans le même esprit, un index spécial met en relation les objectifs visés ou les capacités à exercer avec les exercices prévus et les informations fournies dans les différents livres et moyens à disposition. Ceci permet de donner un entraînement correctif intensif portant sur tel ou tel objectif, un peu comme le font les orthophonistes pour redresser des défauts d'apprentissage. Un index général sert

enfin à rassembler toutes les informations théoriques et pratiques qui sont dispersées dans le livre du fait qu'elles sont rattachées à la préparation des leçons.

Un autre niveau de formation sous forme de séminaire, est mis en place lors de l'introduction de la méthode dans une école. Il est intéressant de noter qu'une séance d'information aux parents est prévue en même temps que le séminaire, et qu'elle s'accompagne de démonstrations, où participent les enfants eux-mêmes.

Conclusions

Rappelons, en conclusion, que si quelques méthodes de lecture ont été présentées schématiquement ci-dessus, c'est uniquement à titre d'exemples de curriculum. Il importe peu de connaître leurs imperfections particulières. Les directions de développement actuelles (vers une meilleure liaison entre les curriculums de mathématiques et de sciences, d'une part et ceux de la lecture et de la langue maternelle d'autre part), ne concernent pas non plus le but que nous visons.

Le problème que nous posent les curriculums est bien plutôt celui du rôle de l'enseignant. Le maître va-t-il continuer à développer lui-même ses méthodes et à fabriquer lui-même ses outils, comme les vieux artisans, amoureux du « tout fait main » ? Va-t-il plutôt suivre une méthode scientifiquement élaborée où la stratégie générale ne dépend plus de lui,

mais seulement l'adaptation intelligent de cette stratégie au cas de chaque enfant ? Sommes-nous même arrivés, comme avec les méthodes de lecture audio-visuelles, à un degré d'individualisation tel que le maître est déchargé de tout apport d'information et n'a plus qu'à apporter un appui individuel et à susciter les activités de groupe ?

Jamais l'enseignement n'a évolué aussi vite, et les perspectives de l'éducation permanente précipitent encore cette transformation. Comme lorsqu'un canot aborde des rapides, ce n'est pas en détournant les yeux des dangers qui approchent que l'on peut maintenir son équilibre. C'est au contraire en les devinant à temps que l'on peut réagir et choisir des passages plus sûrs.

Il nous a semblé qu'une prise de conscience s'imposait chez les enseignants et qu'un débat méritait d'être ouvert rapidement pour ou contre ces curriculums. Plus exactement, si l'on admet que la pratique actuelle constitue déjà une forme de curriculum, il faudrait qu'un accord se dégage sur les caractéristiques souhaitables des curriculums futurs de l'école romande. L'image de l'enseignant qui parle à des élèves qui écoutent doit être renouvelée ; mais par quoi voulons-nous la remplacer ?

Jean Cardinet,

Institut romand de recherche
et documentation pédagogique
(IRD.P).

La page du GREM

Le groupe « Ecole et Vie » des Franques-Montagnes réunit une vingtaine d'enseignants. Il ne fait pas, à proprement parler, partie du mouvement de l'école moderne, mais les idées et les techniques de celle-ci y sont largement répandues (journal et correspondance scolaires, imprimerie, etc.)^{*}.

Les rencontres ont lieu en principe tous les quinze jours dans la classe de l'un ou l'autre des membres. La visite d'une autre classe est toujours l'occasion de découvrir un petit truc, une idée dont on peut ensuite tirer profit pour son enseignement. Il est certain qu'il y a ainsi des tas d'idées valables et originales disséminées à travers le corps enseignant et si tous ceux qui ont de bonnes idées voulaient bien ne pas les garder pour eux seuls et les faire connaître, cela rendrait service à tous et éviterait à beaucoup bien des tâtonnements.

^{*} Il travaille en liaison étroite avec le groupe « Ecole et Vie » d'Ajoie et des contacts fréquents ont lieu avec celui-ci.

Pour rendre plus sympathique l'ambiance des rencontres, chacun apporte un tour de rôle de quoi boire et grignote pendant celles-ci. Parfois, le travail est essentiellement pratique. On met au point ensemble des travaux de géographie, de sciences, de français, de calcul et on le multicopie ensuite pour tout le groupe. L'une des écoles étant bien installée pour les travaux manuels, on y fait des boîtes de rangement pour fichier, des boîtes de raccordement pour écouteurs afin de pouvoir utiliser le magnétophone en classe, et même des étagères, car le groupe dispose d'un important stock de bois.

D'autres fois, le temps de la rencontre (qui finit en principe à sept heures) est utilisé pour une discussion. Les thèmes ne manquent pas : notes et docimologie, programmes, enseignement de telle ou telle branche, sans oublier, bien sûr, cette année, la question de l'année longue ou des trois années courtes, ou celle de la coordination romande. C'est aussi l'occasion

CH. 2. EXEMPLE DE CURRICULUM: LE CURRICULUM ROMAND

Le but du cours est double. Les étudiants du premier cycle doivent s'informer des idées directrices de l'enseignement donné en Suisse Romande. Ceux du deuxième cycle doivent apprendre la façon de mener des études d'évaluation, en analysant de façon critique les recherches effectuées sur les curriculums romands.

Les exposés devront donc se centrer tantôt sur l'une, tantôt sur l'autre de ces deux perspectives. Il va s'agir dans ce chapitre d'une première présentation du curriculum d'ensemble développé pour la Suisse romande et du contexte historique qui a permis son développement.

HISTOIRE DE LA COORDINATION SCOLAIRE ROMANDE

Les débuts de la Coordination

Ce qui a précédé

La Constitution fédérale

les art. 27 et 27 bis de la Constitution fédérale de 1848 stipulent que chaque canton est souverain en matière d'éducation. L'organisation de l'enseignement était donc des plus disparates jusqu'en 1970, année de la signature du Concordat intercantonal suisse, qui explicite quelques points d'accord entre cantons (âge d'entrée à l'école, durée de la scolarité, moment du début de l'année scolaire).

Les demandes des enseignants romands

Déjà en 1898, mais surtout en 1962, au Congrès de Bienne, la Société Pédagogique Romande (SPR), syndicat des instituteurs romands, demande d'harmoniser les programmes et les structures pour créer une école semblable dans toute la Suisse romande. Les enseignants recevraient aussi une formation uniforme. Cette coordination étroite serait obtenue par voie concordataire.

La réponse des milieux politiques

L'expansion économique que connaît la Suisse à cette époque permet toutes les audaces: le Conseiller d'Etat vaudois, Pierre Oguey, répond positivement. Le chanoine Barbey à Fribourg pousse à la roue en prenant aussi le parti des jeunes qui changent de canton et sont perturbés dans leurs études.

En 1963, la CDIP, lors de sa rencontre annuelle, crée trois commissions relatives à la coordination des études et des moyens d'enseignement.

En 1965, les Chefs de Services des différents cantons décident de se rencontrer régulièrement.

La phase préparatoire

Les années suivantes, la SPR maintient sa pression et prépare des propositions concrètes, avec des projets de programmes et des suggestions d'institutions, par exemple un Centre de recherches pédagogiques romand.

Le lancement

En 1967 les autorités scolaires prennent des décisions effectives: la Commission Interdépartementale Romande de Coordination de

l'Enseignement (CIRCE) est créée pour étudier de nouveaux programmes.

Cette Commission décide immédiatement de faire du neuf en s'inspirant des recherches pédagogiques les plus récentes, plutôt que de chercher des compromis entre les pratiques cantonales existantes.

En 1968 et 1969, les deux premiers fonctionnaires romands sont nommés, chargés de faire progresser la coordination (André Neuenschwander et Jean Cavadini); en 1970, c'est l'Institut Romand de Recherches et de Documentation Pédagogiques (IRDP) qui est créé à Neuchâtel.

C'est la même année qu'est signé le Concordat suisse à Montreux, le 29 octobre 1970.

Les nouveaux programmes

CIRCE 1 remet les projets de programmes pour les 4 premières années de la scolarité, le 15 décembre 1972 à la Conférence des Chefs de Départements et Directeurs de l'Instruction Publique de la Suisse Romande et du Tessin (CDIP SR+TI).

Le rôle politique de l'évaluation au départ

Les premières résistances

Certaines oppositions (fédéralistes) sont motivées par la crainte de la centralisation, même si la coordination romande est présentée comme une mesure destinée à prévenir l'intervention de la Confédération en matière d'enseignement.

D'autres craintes sont économiques, devant les projets ambitieux des rénovateurs.

Mais surtout l'audace des transformations projetées inquiète l'opinion publique, qui voit déjà le niveau baisser...

Les plans d'études dits expérimentaux

Pour apaiser les craintes, les autorités déclarent qu'elles sont prêtes à modifier ces projets sur la base de l'expérience. Jean-Pierre Pradervand, en adoptant les programmes de CIRCE, déclarait: "Un plan d'études n'est pas immuable, il doit subir l'épreuve de l'expérience qui amènera certaines corrections."

L'IRDP reçoit officiellement le mandat d'évaluer ces programmes en proposant les adaptations nécessaires.

Nature de l'évaluation envisagée

On peut noter cependant qu'il n'a jamais été question pour personne de revenir en arrière. L'évaluation envisagée devait être "formative", elle devait donner des indications sur les corrections à apporter pour mieux atteindre les nouveaux objectifs. Il ne s'agissait pas d'évaluer ces objectifs eux-mêmes, qui étaient devenus des choix politiques définitifs.

Suite de l'histoire de la coordination

L'introduction progressive des disciplines

Les changements envisagés paraissaient d'une telle ampleur que les Chefs de Services décidèrent dès le départ de fractionner la tâche (surtout celle de recyclage du corps enseignant). Ils établirent un calendrier d'introduction qui prévoyait l'année de la mise en

vigueur de chaque nouveau programme dans chaque degré de chaque canton.

En raison de divers incidents de parcours, (rejet de certains moyens d'enseignement, moratoires imposés par des Grands Conseils, etc.) l'ensemble s'est étalé sur environ 20 ans.

La première branche rénovée fut la mathématique, puis un certain nombre de disciplines secondaires. puis le français, puis après d'autres branches secondaires encore, l'allemand.

Les difficultés ultérieures de la Coordination

Des difficultés externes

Des oppositions sont apparues à propos de chaque programme nouveau. La mathématique moderne a été critiquée, comme trop abstraite, par les professeurs de l'EPFL et les milieux professionnels.

Le nouvel enseignement du français, qui abandonnait la grammaire traditionnelle à finalité orthographique au profit de la linguistique, a soulevé des polémiques dans tous les cantons. Il a failli plusieurs fois être condamné par les milieux politiques, qui ont en tout cas retardé son introduction de plusieurs années.

L'option communicative et orale de l'enseignement de l'allemand a rencontré la résistance des professeurs de l'enseignement secondaire supérieur, attachés à la valeur culturelle de l'étude de la littérature allemande.

Des difficultés internes au système scolaire

La plus grande résistance est venue cependant des milieux de l'enseignement secondaire, menacés par les réformes de structure demandées par la SPR et par les changements de méthodes qu'imposerait une coordination romande.

Ils ont pu bloquer jusqu'ici toute coordination véritable dans les trois dernières années de la scolarité obligatoire, en exigeant que les "programmes-cadres" ne soient pas définis par année et en refusant d'admettre des moyens d'enseignement communs.

De fait, avec des structures scolaires aussi différentes, des taux de sélection allant du simple au double, et des horaires de branche allant également du simple au double, on voit mal ce qui pourrait être commun d'un canton à un autre.

Le rôle politique de l'évaluation par la suite

L'appel aux faits plutôt qu'aux arguments a priori

Dans cette atmosphère conflictuelle, la recherche pédagogique a été appelée à jouer un rôle important. Elle a permis de faire des constats rassurants, mais aussi d'introduire certaines des modifications souhaitées par les enseignants. Elle a ainsi réduit le caractère passionnel des débats et contribué à une évolution plus progressive de l'enseignement.

Les exigences méthodologiques de cette évaluation

Pour être crédibles, les rapports des chercheurs devaient faire la preuve de leur validité, c'est-à-dire être reconnus comme objectifs (faisant l'accord des observateurs) par d'autres chercheurs.

Ceci supposait donc l'emploi de méthodes d'observation éprouvées; mais il fallait aussi utiliser des démarches d'interprétation "participatives", où les valeurs, les vœux et les craintes des divers partenaires soient pris en compte dans les conclusions. La frontière entre évaluation et gestion est donc imprécise, dans cette forme de recherche qui collabore aussi à l'action.

Des règles de conduite (d'éthique professionnelle) doivent être suivies avec beaucoup d'attention si l'on veut éviter le danger de manipulation de l'opinion publique. Les risques sont très grands de voir la recherche utilisée contre tel ou tel adversaire, plutôt que comme un phare qui éclaire l'ensemble de la situation.

L'organisation de ce cours

En étudiant l'histoire des évaluations effectuées depuis 20 ans par les services de recherche pédagogique romands, les étudiants vont pouvoir se familiariser avec les problèmes que l'on rencontre habituellement et les solutions possibles.

La discussion des exemples présentés devra faire appel au sens critique des participants et justifier l'importance d'une explicitation des principes à suivre, qui devrait être réalisée au semestre d'été.

CONTENU DES PLANS D'ÉTUDES

C'est en présentant le curriculum de chaque branche que l'on analysera en détail ses objectifs, par rapport auxquels les évaluations ont pu être conduites.

Mais il faut au départ analyser les objectifs éducatifs généraux qui ont implicitement guidé le choix de ces curriculums particuliers. Le plan d'études romand est en effet remarquablement cohérent dans son approche pédagogique. Il aurait été regrettable de ne pas en tenir compte dans les adaptations proposées par la suite.

Raisons de l'absence d'objectifs généraux

On ne trouvera pas, pourtant, de formulation explicite des objectifs éducatifs communs, en tête des plans d'études.

Une première raison tient à la méfiance du caractère suisse vis-à-vis de tout ce qui paraît idéologique et théorique. Le président de CIRCE, Roger Nussbaum, défend ainsi son attitude pragmatique: "CIRCE avait le choix entre deux attitudes: manipuler des éléments de doctrines, peser des principes, esquisser de fabuleux paysages pédagogiques ou choisir une démarche, moins exaltante peut-être, qui tienne compte du poids des données actuelles. Cette manière plus réaliste de voir les choses a été retenue."

Une seconde raison tient à la réalité politique de la Suisse romande, où les Départements de l'Instruction Publique reviennent à des partis aux idéologies très différentes d'un canton à l'autre. Durant la première phase de la Coordination, le GROS (Groupe de Réflexion sur les Objectifs et les Structures) avait réussi à déterminer des objectifs éducatifs communs pour la Suisse romande, mais la CDIP SR+TI refusa de les approuver, car cela aurait signifié qu'un accord politique était possible entre ces partis sur une conception commune de l'éducation. Il a donc fallu que les partenaires de la Coordination

réalisent des choix sans avoir de référence théorique pour les fonder, contrairement aux lois sur l'instruction publique cantonales, qui proposent toutes des buts à l'éducation.

Objectifs généraux implicites

Evidemment, on ne peut agir sans se référer à des valeurs; il existe donc bien des objectifs généraux derrière les plans CIRCE, même s'ils doivent rester informulés. Sans doute le développement de l'autonomie des enfants a-t-il été au centre des préoccupations des éducateurs qui ont rédigé les plans romands: apprendre à résoudre des problèmes, à raisonner logiquement, à analyser des idées de façon critique, à s'exprimer facilement et en se faisant comprendre, à communiquer en dépit des barrières linguistiques.

L'approche pédagogique choisie, surtout, est très homogène: elle s'inspire directement des conceptions piagétienne, qui demandent de faciliter l'activité de l'enfant, pour qu'il construise ses instruments de pensée au contact de la réalité physique et sociale. Les activités de recherche, notamment en groupes, sont encouragées, aussi bien en mathématique qu'en français et en allemand. La mémorisation doit toujours s'appuyer sur une phase préalable de découverte et de construction personnelle des concepts, (de sensibilisation, disent les plans d'études).

Objectifs de branches

Des buts généraux sont formulés dans les trois plans d'études (CIRCE 1 pour les degrés 1 à 4, CIRCE 2 pour les degrés 5 et 6, CIRCE 3 pour la fin de la scolarité obligatoire) pour chacune des branches. Dans une reformulation ultérieure, celle du Groupe Romand d'Aménagement des Programmes (GRAP), le texte propose pour chaque branche deux types d'objectifs, d'une part des compétences globales, comme savoir raisonner logiquement, à développer dans des situations aussi riches de signification que possible, d'autre part des savoir faire spécifiques, à exercer en classe jusqu'à leur maîtrise, qui doit être acquise à un âge déterminé.

Le piège des programmes traditionnels, qui indiquent des contenus, mais ne précisent pas ce que les élèves doivent être capables de faire à leur propos, semble ainsi avoir été évité.

CONSTRUCTION ET INTRODUCTION DES CURRICULUMS ROMANDS

La stratégie d'introduction

Les changements de plan d'études, dans la majorité des pays, sont décrétés au niveau du Ministère, et les enseignants sont sensés s'y adapter. (Des recherches effectuées en Belgique ont montré qu'en réalité les changements demandaient des années.)

Le Tessin a changé son plan d'études primaires il y a quelques années, en se contentant de préciser les nouveaux objectifs et en apportant une formation générale aux enseignants par l'intermédiaire des inspecteurs.

La Suisse Romande, au contraire, a opté pour une démarche plus lourde, s'appuyant sur de nouveaux moyens d'enseignement très

détaillés, sur une formation approfondie des instituteurs, et encore sur des recherches d'accompagnement destinées à mieux assurer les chances de réussite de l'opération.

La rédaction des moyens d'enseignements

Pour chaque branche et chaque degré, plusieurs spécialistes ont été nommés par la Commission Romande des Moyens d'Enseignement (COROME) avec mission de préparer des méthodologies (à l'intention des enseignants), des manuels et des livres d'exercices (pour les élèves), des moyens d'enseignement complémentaires (jeux didactiques, bandes audio et vidéo, fichiers, mémentos, etc.), et depuis peu des moyens d'évaluation et de rattrapage.

Leur expérimentation et leur mise au point

Ces productions ont toujours été soumises à des Commissions d'examen romandes, qui ont demandé des adaptations diverses.

Depuis une dizaine d'années, ces manuscrits donnent lieu au tirage d'une forme préalable, qui est expérimentée dans des classes. Les remarques des enseignants et les observations des chercheurs servent de base pour la préparation de la forme finalement diffusée dans l'ensemble des classes.

La formation des enseignants

Lorsque les moyens d'enseignement et les méthodologies sont prêtes, les enseignants reçoivent une formation spéciale pour être en mesure de les utiliser dans l'esprit de la réforme.

Ces recyclages ont duré deux ans pour la mathématique, plus d'un an pour le français et l'allemand, durées variables d'ailleurs selon les cantons, les degrés, les crédits disponibles, etc.

Les Départements utilisaient la démarche de la "boule de neige": quelques spécialistes universitaires formaient dans chaque canton des animateurs-trices (pour un degré donné), qui se chargeaient à leur tour de former des collègues l'année suivante, en prenant en charge une dizaine de personnes par groupe.

L'évaluation des programmes et des moyens d'enseignement

La recherche pédagogique n'est intervenue, dans les premières années de la Coordination, (introduction de la mathématique) qu'après que l'innovation ait été diffusée et appliquée au moins un an. L'idée était de laisser passer les moments de crise inévitables lors d'un changement de méthode, pour n'observer les résultats que dans des situations normales.

Ultérieurement (introduction du français), c'est la façon même dont le nouvel enseignement était présenté aux enseignants et introduit dans les classes qui faisait l'objet de recherches. L'idée était cette fois de faciliter la réussite de l'opération en vérifiant comment elle était lancée et en proposant, si nécessaire, des corrections immédiates.

Leurs remaniements

C'étaient les moyens d'enseignement qui étaient les plus faciles à modifier, mais à plus long terme les méthodologies, et même les plans d'études, ont pu faire l'objet de propositions d'adaptation.

A cause du temps nécessaire à la collecte et à l'analyse des observations, des adaptations immédiates se sont révélées difficiles. Les conclusions ont pu servir néanmoins pour les degrés suivants, et aussi pour les cantons qui introduisaient le nouvel enseignement après les autres.

ASPECTS INSUFFISAMMENT DÉVELOPPÉS DE FAÇON GÉNÉRALE

Pour terminer cet aperçu préalable de l'introduction des nouveaux curriculums, on peut voir ce qui leur a manqué en général.

L'évaluation des élèves

Les auteurs des méthodologies et des moyens d'enseignement ont fait comme si l'évaluation était du domaine réservé à chaque maître. Il est vrai qu'il est impossible d'avoir les mêmes exigences d'un milieu à un autre. Une évaluation formative, par contre, pourrait s'appuyer sur des exercices diagnostiques identiques.

L'individualisation

La suite logique d'une évaluation formative est d'adapter les démarches d'apprentissage aux insuffisances observées.

Faute de moyens d'évaluation, rien n'a non plus été prévu dans les moyens d'enseignement habituels. Seules les classes à degrés multiples ont reçu des fichiers permettant de donner des exercices adaptés à des élèves se situant à des niveaux d'avancement différents.

Les problèmes des enfants migrants

Ce n'est qu'en 1994 que seront distribués aux enseignants certains moyens de faire face aux problèmes que leur pose l'arrivée d'enfants ayant d'autres langues et d'autres cultures.

CH. 3. THÉORIE DE L'ÉVALUATION DE CURRICULUM

DÉFINITIONS

Définition de l'évaluation

Il en existe de nombreuses, mettant l'accent tantôt davantage sur l'observation, tantôt sur les critères de référence, tantôt sur l'acte de jugement, ou ses conséquences. Celle de Jean-Marie De Ketele (Observer pour éduquer, Berne: Peter Lang, 1980) est assez complète et nuancée:

"Évaluer consiste à recueillir un ensemble d'informations reconnues comme suffisamment pertinentes, valides et fiables et à examiner le degré d'adéquation entre cet ensemble d'informations et un ensemble de critères jugés suffisamment adéquats aux objectifs fixés au départ ou ajustés en cours de route, en vue de fonder une prise de décision."

Définition de l'évaluation de curriculums

C'est l'étude systématique du fonctionnement d'un curriculum et sa comparaison à des standards de référence, visant à guider les décisions nécessaires au bon fonctionnement de l'école.

Stufflebeam (dans Systematic Evaluation, p. 159) donne une définition très élaborée: "L'évaluation d'un curriculum est le procédé par lequel on sélectionne, recueille et analyse des informations, descriptives et appréciatives, concernant la valeur des objectifs, de l'organisation, de la mise en oeuvre, et des effets de ce curriculum, dans le but de fonder des décisions, de justifier son emploi, et de faciliter la compréhension des phénomènes correspondants."

APPROCHE HISTORIQUE

Les étapes du développement de l'évaluation de curriculum

Marc-André Nadeau, dans le 1er ch. de "L'évaluation de programme", donne une histoire détaillée de cette pratique, telle qu'elle a été développée aux Etats-Unis. Les points principaux en sont repris ci-dessous:

Avant 1900:

On cherche à contrôler le bon fonctionnement des écoles par des observations sur place.

1900-1930:

On quantifie l'efficacité de l'enseignement en faisant passer des tests de connaissances aux élèves.

1930-1945:

Ralph Tyler conceptualise le curriculum (l'ensemble des expériences qui doivent aider les élèves à atteindre certains objectifs). L'évaluation de curriculum revient, pour lui, à déterminer le degré d'atteinte des objectifs visés.

Il applique ces principes à l'étude des effets de l'École Active (Progressive Education). Il opérationnalise les objectifs éducatifs très globaux de ces écoles (affectifs autant que cognitifs) et montre comment comparer les résultats obtenus aux résultats attendus. Sa méthode a les avantages suivants: elle ne nécessite pas de groupe de contrôle; elle se centre sur les résultats effectifs et non sur les moyens, (comme les cours donnés ou les programmes); elle utilise un éventail très large de mesures de résultats, par opposition aux tests de connaissances antérieurs.

1946-1957:

Benjamin Bloom et ses collègues développent leur Taxonomie des objectifs éducatifs, dans le domaine cognitif, puis dans le domaine affectif.

L'opérationnalisation des objectifs progresse donc et l'on fait confiance aux tests pour donner des points de référence précis, permettant de guider l'apprentissage individuel et la mise au point des curriculums.

1958-1972:

A la suite du lancement du Sputnik, les Etats-Unis cherchent à mettre au point de nouveaux curriculums de mathématique et de sciences. Leur évaluation est également demandée, ce qui déclenche un grand nombre d'études d'évaluation, se référant aux résultats soit des groupes contrôles, soit aux objectifs prédéfinis.

Mais Lee Cronbach publie en 1963 "Course improvement through evaluation", montrant que ces études ne servent presque à rien, parce qu'elles comparent des curriculums ayant des objectifs différents et souvent se contentent d'expérimenter une forme particulière d'un cours particulier. Elles négligent les problèmes importants qui permettraient, seuls, de faire progresser l'éducation, en faisant mieux comprendre les facteurs de l'apprentissage. Il serait plus intéressant, par exemple, d'analyser les taux de réussite question par question que de comparer les moyennes du groupe expérimental à celles du groupe contrôle. L'évaluation devrait être repensée comme un processus de recueil d'informations utiles pour le développement de nouveaux curriculums à l'avenir, au lieu d'être centrée sur les curriculums déjà terminés.

Michael Scriven se trouve en accord avec Cronbach sur l'idée que les effets des curriculums étudiés doivent être examinés sur toute une série de dimensions, d'abord parce qu'ils peuvent avoir des effets indirects imprévus, et ensuite parce que c'est la seule façon de pouvoir comparer des enseignements qui visent des objectifs différents. Il demande donc une "évaluation dégagée des objectifs" (goal free).

En 1965 Kennedy lance "la Guerre à la Pauvreté", qui comporte beaucoup de programmes éducatifs, mais qui exige leur évaluation scientifique. Ces évaluations sont généralement décevantes, parce que les tests pédagogiques à disposition mesurent essentiellement les mêmes facteurs que les tests d'intelligence et sont donc peu affectés par les expériences éducatives. De plus, les études comparatives ont

peu de sens, parce que les conditions d'une expérimentation rigoureuse ne sont presque jamais réunies. Les études se référant aux objectifs pèchent de leur côté par le fait qu'on ne sait pas déterminer les résultats minima à exiger. De toute façon, ces comparaisons ne prouvent rien dans la mesure où l'on ne sait jamais de quelle façon le "traitement" prévu a été appliqué par les enseignants et reçu par les élèves.

Cette époque stimule donc la réflexion méthodologique. Les chercheurs demandent en particulier de: 1) évaluer les objectifs éducatifs eux-mêmes 2) étudier le contexte pour déterminer les besoins 3) noter en détail comment le programme à évaluer a été introduit et appliqué 4) étudier les effets imprévus aussi bien que prévus 5) mettre les résultats en rapport avec les coûts des différentes méthodes existantes.

1973-1993

Les personnes qui s'occupent d'évaluation se trouvent une identité, créent des revues scientifiques, se donnent des normes de qualité (standards) et une déontologie, se regroupent dans des sociétés nationales et internationales... Mais de nombreux problèmes demeurent encore aujourd'hui.

Différenciation progressive de l'évaluation et de la recherche

L'histoire de l'évaluation de programme peut suivre un autre fil rouge: celui de la constitution progressive d'un champ conceptuel et d'une pratique distincts de celui de la recherche scientifique habituelle.

L'évaluation comme forme de recherche

Au départ, au siècle passé et jusqu'aux travaux de Tyler, l'évaluation de curriculum est conçue comme une recherche scientifique portant sur les performances obtenues des élèves selon les conditions d'apprentissage qui leur sont fournies.

La méthode expérimentale permet de s'assurer que les effets observés proviennent uniquement du traitement pédagogique appliqué. On peut alors légitimement porter un jugement sur l'efficacité de ce traitement.

L'évaluation comme activité spécifique

Pourtant l'évaluation d'un curriculum dépend de nombreux facteurs supplémentaires: pertinence des objectifs visés, façon dont s'est déroulé le programme, coût de son introduction, etc. Une recherche centrée sur une décision suit une logique différente de celle d'une recherche centrée sur une conclusion.

Marc-André Nadeau développe, dans son deuxième chapitre, les points sur lesquels recherche et évaluation se différencient. Son Tableau 1, qui résume ces différences, est reproduit ci-dessous.

Tableau 3.1.
DIFFÉRENCES FONDAMENTALES ENTRE LA RECHERCHE ET L'ÉVALUATION

Éléments	Recherche	Évaluation
Motivation du chercheur	Curiosité scientifique	Solution d'un problème

Objectifs de l'étude	Conclusions	Décisions
Rôle de l'explication	Recherche d'explications	Détermination de valeur
Liberté du chercheur	Grande autonomie	Faible autonomie
Utilité sociale	Indirecte	Directe
Généralisabilité	Grande	Faible
Contrôle expérimental	Essentiel	Possible
Critères pour juger	Validité interne et ext.	Isomorphisme crédibilité
Base disciplinaire	Unidisciplinaire	Multidisciplinaire

Conditions d'émergence des évaluations de formations

Jean-Marie Barbier, dans son ouvrage "L'évaluation en formation", Ch. 1, relève le parallélisme étroit entre les conditions d'apparition des évaluations concernant d'un côté les processus de production industrielle et de l'autre les processus de formation des personnes. Il s'agit dans l'un et l'autre cas d'efforts de la part de responsables pour rationaliser la gestion d'un certain travail, en faisant appel à la participation des intéressés.

Ainsi, la pédagogie orientée vers des objectifs est à rapprocher de la direction par objectifs; la première se substitue à l'inculcation de connaissances, comme la seconde au mode de commandement autoritaire.

C'est la même idéologie qui anime les travaux de l'OCDE sur les indicateurs de fonctionnement des systèmes scolaires, ou les enquêtes internationales de l'International Educational Assessment (IEA). Toute gestion efficace repose sur une évaluation des résultats.

APPROCHE PRESCRIPTIVE

Toute une série de modèles pour guider une recherche évaluative ont été proposés depuis 20 ans. Marc-André Nadeau les présente dans ses chapitres 3 à 5. On les retrouvera ci-dessous, pour l'essentiel, mais organisés selon d'autres catégories, qui devraient permettre de choisir plus systématiquement le modèle le plus approprié à chaque cas.

Épistémologie objectiviste

Ce qui caractérise les méthodes de ce type (celles que Nadeau appelle "formalistes" p. 66), c'est, dit-il, qu'elles s'inspirent de l'analyse de système, en considérant le système scolaire comme composé de trois éléments: intrant, processus et extrant. Elles sont formalistes dans le sens qu'elles précisent dès le point de départ l'orientation qu'elles veulent prendre et les informations qu'elles veulent recueillir, généralement à l'aide d'instruments standardisés qui fournissent des informations quantitatives. Elles formalisent enfin leur processus de jugement et de décision et l'explicitent dans un rapport écrit.

Plus profondément, c'est qu'elles admettent qu'une connaissance objective est possible, que des évaluateurs peuvent s'entendre sur des démarches systématiques pour atteindre la vérité, en tout cas un

accord entre tous les observateurs de bonne foi. Elles fonctionnent donc avec la même épistémologie que les sciences de la nature.

Utilité commune

Le premier sous-groupe admet aussi qu'il est possible de s'entendre sur une échelle de valeurs à l'intérieur de notre société, de façon à arriver à une mesure commune de l'utilité de chaque décision envisageable. Le rôle de l'évaluateur serait de parvenir à estimer ces utilités par des démarches contrôlables.

Orientation théorique: vers la connaissance

La troisième dimension de classification se réfère à l'opposition mentionnée plus haut entre recherche pour savoir et recherche pour agir.

Évaluation expérimentale (comparative)

Comme il a été dit, la conception initiale de la démarche évaluative au début du siècle était celle de la pédagogie expérimentale. Ce modèle est défendu encore par des sociologues purs et durs comme Rossi, Freeman et Wright (1979). Par des démarches de laboratoire, on veut découper la réalité pour isoler l'effet causal propre du traitement pédagogique envisagé. Selon les standards de la méthodologie statistique la plus orthodoxe, il faut, pour que le test statistique soit valide, comparer les progrès du groupe expérimental à ceux du groupe contrôle, (ou des groupes contrôles). S'il l'effet du traitement se révèle suffisamment positif pour se distinguer des fluctuations aléatoires, on peut recommander l'adoption de ce traitement.

Évaluation opérationnelle (non comparative)

Ralph Tyler a introduit un premier élargissement à ce modèle expérimental. Il fut le premier à défendre l'idée qu'il suffisait pour valider un curriculum de prouver qu'il permettait d'atteindre les objectifs éducatifs visés. La comparaison à un groupe contrôle ne faisait que compliquer la situation. C'était ensuite aux décideurs de prendre en compte les autres aspects économiques et politiques de la gestion de l'école.

Orientation pratique: vers la décision

C'est au milieu de ce siècle qu'a été développée la théorie de la décision, comme cadre rationnel englobant, valable pour la statistique, comme pour l'ensemble des activités pratiques. Des chercheurs comme Lee Cronbach ont montré l'élargissement que ce cadre de référence plus large apportait à la méthodologie des sciences de l'éducation.

Évaluation systémique (comparative)

Dans la réalité, les décideurs ne peuvent pas isoler un facteur des autres. Même en l'absence d'information, ils doivent prendre un parti. Ils doivent choisir une option en sachant qu'elle aura des effets indirects difficilement contrôlables. La politique est l'art du possible: elle recherche le moindre mal, mais ne peut éviter le risque. Le choix doit donc s'appuyer sur des indices, de simples probabilités souvent.

C'est pourquoi des analyses systémiques sont menées par des chercheurs comme Gilbert De Landsheere, l'un des promoteurs de

l'IEA. Les résultats obtenus par les différents systèmes d'enseignement du monde sont analysés et comparés, et certaines conclusions en ressortent (influence de l'effectif de la classe, de l'âge de la sélection, etc.). Jamais les conditions d'une preuve expérimentale ne sont satisfaites, mais les tendances observées, souvent à l'aide de corrections statistiques, peuvent servir d'arguments pour proposer telle ou telle politique.

L'OCDE a demandé au CERI de développer un ensemble d'indicateurs du fonctionnement des systèmes scolaires, qui faciliterait ces comparaisons et permettrait de gérer l'instruction publique comme l'économie nationale. L'entreprise est risquée socialement et politiquement, même si elle est scientifiquement intéressante.

Évaluation comptable (non comparative)

Les indicateurs de performance peuvent être utilisés simplement pour "demander des comptes" au système scolaire.

C'est ce que souhaitait un mouvement politique aux USA (pour la responsabilisation des services éducatifs = accountability) mécontent du fonctionnement des écoles. Ce mouvement réclamait la mise en place systématique d'une pédagogie de maîtrise (avec contrôle continu et remédiation), demandait que les enseignants tiennent la comptabilité de leurs résultats, et voulait même lier leur salaire à leur efficacité mesurée.

En liaison avec cette nécessité de "rendre compte", beaucoup d'états américains ont exigé que les élèves passent en fin de scolarité des "tests de compétences minimales" qui permettent de contrôler à la fois les élèves, les enseignants et les écoles.

La recherche pédagogique a beaucoup à faire pour développer des tests sommatifs de ce genre qui rendent compte correctement des objectifs majeurs de l'éducation de base et du fonctionnement du système scolaire.

Utilités particulières

Dans la majorité des cas précédents, l'utilité supposée commune est en réalité l'utilité estimée par l'instance de décision. Même si cette dernière est démocratiquement élue, elle reste la représentante d'une partie seulement de la population, la fraction majoritaire.

La démocratie actuelle tend le plus possible à tenir compte aussi des avis minoritaires, souvent également légitimes et porteurs de progrès.

Les évaluations doivent donc se faire par rapport à des cadres de référence multiples, par exemple ceux des parents, des enseignants, des spécialistes de la discipline, etc., dont les valeurs sont nécessairement différentes.

Dans ce cas, la comparaison des utilités par rapport à une échelle de valeur commune est impossible. C'est pourquoi seules des approches non comparatives sont à considérer dans la liste ci-dessous.

Orientation théorique: vers la connaissance**Évaluation dégage des objectifs (non comparative)**

Michael Scriven développa cette approche pour éviter le biais qui résulte de la considération exclusive des objectifs annoncés pour un programme: il est préférable, dit-il, d'examiner les effets, quels qu'ils soient, même s'il s'agit d'effets imprévus.

Naturellement, il n'est pas possible de tout mesurer et c'est l'évaluateur qui doit faire des hypothèses sur les effets probables qu'il va observer, sur la base du curriculum appliqué: c'est donc un des risques de la méthode.

Ensuite, un résultat observé n'a de sens que si on peut démontrer sa liaison avec le traitement utilisé. Une recherche explicative est donc nécessaire, pour laquelle aucune démarche standard n'est offerte. C'est un deuxième risque.

Enfin Scriven souligne que l'échelle de valeur à prendre en compte en priorité est celle des élèves et des enseignants, les vrais "consommateurs" du curriculum, et non celle des auteurs ou de l'administration. Le fait de ne pas prendre le même point de vue que ceux qui financent l'étude peut aussi faire problème. Il reste néanmoins utile de montrer que ces utilités sont diverses.

Orientation pratique: vers la décision**Évaluation décisionnelle (non comparative)**

Daniel Stufflebeam a été un des premiers à présenter l'évaluation de curriculum comme un processus de feedback destiné à aider les responsables éducatifs à prendre leurs décisions. L'évaluateur doit cerner les décisions à prendre et fournir les éléments permettant de faire des choix rationnels. C'est dans cette perspective qu'il doit sélectionner les informations pertinentes et les analyser.

Stufflebeam analyse le problème en 4 étapes: l'évaluation du contexte, de l'intrant, du processus et du produit, chacune comportant des décisions particulières à prendre.

- La première étape doit aboutir à la spécification des buts et des objectifs du curriculum. On y parvient sur la base d'une analyse des besoins, c'est-à-dire en examinant dans le contexte les problèmes nécessitant une amélioration de la formation.
- L'évaluation de l'intrant consiste à identifier et à comparer les différents moyens d'obtenir les résultats visés dans le contexte en question. La décision correspondante est le plan d'action à appliquer.
- L'évaluation du processus intervient quand le programme commence à être appliqué. Elle doit préparer les décisions qui amélioreront l'introduction du programme et ses procédures d'application.
- L'évaluation de produit mesure les effets du programme et les interprète en les comparant aux objectifs, compte tenu du contexte, de l'intrant et des processus d'application. Elle doit conduire aux décisions de rejet, de révision ou de diffusion du curriculum.

Critique des méthodes formalistes

- elles fournissent des réponses aux questions des chercheurs, mais pas aux questions des enseignants ou des autorités.

- les critères servant à déterminer si les objectifs sont atteints sont rarement satisfaisants, la validité des tests pour mesurer le résultat d'une éducation, par exemple, pouvant être contestée.
- les données recueillies sont choisies à l'avance et risquent de laisser de côté les vrais problèmes.
- les questions traitées sont en nombre limité.
- les préoccupations des autorités reçoivent plus d'attention que celles des enseignants ou des parents.
- les chercheurs extérieurs perçoivent mal les préoccupations du milieu où ils interviennent et les enjeux de leur intervention.
- les chercheurs se préoccupent plus d'éviter les critiques méthodologiques de leurs pairs que de répondre aux questions pratiques qui leur sont posées.

Épistémologie subjectiviste

Dans le domaine social, c'est une hypothèse risquée que d'admettre, comme les modèles précédents, que les observateurs de bonne foi doivent arriver aux mêmes conclusions. Il est bien plus probable que chacun privilégiera inconsciemment tel ou tel aspect de la réalité. Beaucoup d'aspects de notre expérience sont même non verbaux, intuitifs, subjectifs, non comparables.

D'autres modèles d'évaluation partent donc de l'hypothèse que les réalités sont multiples, même si les nécessités de la vie sociale obligent à arriver à certains compromis permettant des activités cohérentes. Marc-André Nadeau les appelle "naturalistes".

Ce qui caractérise ces méthodes, c'est une démarche inductive, qui part d'une observation sans hypothèse a priori pour faire émerger petit à petit des tentatives de généralisation, que l'on confirme progressivement. L'instrument principal est l'intuition de l'observateur, bien que ce dernier tende à structurer et recouper ses impressions par divers moyens. Le souci principal est celui de la pertinence: dégager les dimensions les plus significatives qui rendent compte de la vision subjective des différentes audiences.

Il va de soi que si les réalités subjectives sont différentes, les valeurs sont aussi personnelles. C'est pourquoi le seul cas possible dans cette catégorie est celui d'échelles d'utilité particulières.

Utilités particulières

L'évaluateur doit, dans cette hypothèse, s'adresser séparément à ses différentes audiences: autorités politiques, administrateurs, enseignants, parents, élèves...

Son rôle sera de faire circuler entre eux les informations concernant leurs perceptions et leurs attitudes respectives, pour faciliter leur dialogue.

Orientation théorique: vers la connaissance

Évaluation de connaisseur (non comparative)

C'est Eisner (1977) qui a forgé ce terme, en se référant à la pratique des critiques d'art, de cinéma, de vins, etc., qui font constamment des évaluations reconnues comme valables, sans pourtant se référer à une méthodologie scientifique. Des experts en éducation peuvent

aussi apprécier des aspects subtils des curriculums actuels ou en préparation.

Le rapport que rédigent les critiques contient habituellement trois parties: la description, l'interprétation et l'évaluation.

La description essaye de faire apparaître les traits les plus marquants des démarches éducatives proposées, et ceux qui pourraient échapper à un observateur superficiel. Le critique utilise alors un style imagé pour transmettre de la façon la plus vivante possible l'aspect subjectif de ces expériences.

L'interprétation essaye d'organiser en un tout cohérent les traits relevés dans la description. Elle peut se faire en se référant à un contexte historique ou à des théories issues des sciences sociales, si la théorie des curriculums se révèle insuffisante. Le style est alors discursif et linéaire.

L'évaluation enfin est à développer explicitement (plutôt qu'implicitement comme dans les parties précédentes) en se référant à des critères précis. Ces derniers pourraient être personnels, mais doivent être explicités, pour permettre au lecteur de les juger lui-même, et aussi pour développer dans le public une capacité d'évaluation plus autonome.

Autant que possible, l'évaluation devrait se faire dans le cadre d'une situation d'enseignement réelle, bien que des experts puissent donner des avis pertinents sur des matériels d'enseignement ou des projets de cours.

La généralisabilité des appréciations portées de cette façon repose sur la possibilité de créer chez différents lecteurs des images anticipatrices qui correspondent à leur expérience ultérieure réelle en situation d'enseignement. Si cette généralisabilité est suffisante, cette forme d'évaluation pourrait devenir tout à fait crédible.

Orientation pratique: vers la décision

Évaluation répondante (non comparative)

Robert Stake a évolué au cours de sa carrière dans la direction d'une plus grande souplesse et a abouti en 1973 à une forme d'évaluation qu'il appelle en anglais "responsive", c'est-à-dire sensible, capable de percevoir les multiples dimensions de la réalité et d'y réagir en poussant son investigation dans toutes les avenues suggérées par ses partenaires. (La traduction par "répondante" est très insatisfaisante, mais "réagissant bien" n'est pas meilleur...)

Les traits caractéristiques de son approche peuvent être mis en parallèle avec ceux d'une approche formaliste, dans le tableau suivant:

Tableau 3.2.

Ce qui distingue une évaluation répondante d'une évaluation formaliste

Trait distinctif	Évaluation formaliste	Évaluation répondante
------------------	-----------------------	-----------------------

Intention	Déterminer dans quelle mesure les buts ont été atteints	Aider le client à apercevoir et à traiter les faiblesses du programme
Services offerts	Fournir les informations prévues dans le contrat de départ	Répondre en cours d'étude aux demandes des diverses audiences
Contrat	Au départ, les obligations respectives sont définies aussi précisément que possible	Les intentions et les procédures sont esquissées au départ, puis précisées par la suite
Orientation principale	Objectifs et indicateurs	Problèmes, événements
Plan de l'étude	Prédéterminé	Émergent
Méthodologie	De laboratoire: intervenir et observer les effets	Observer et interpréter ce que font les gens
Techniques préférées	Plan expérimental, hypothèses, objectifs comportementaux, tests, statistiques, rapports	Études de cas, interviews, auditions, observations, débats contradictoires, comptes rendus
Communication entre l'évaluateur et le client	Formelle et irrégulière	Libre et continue
Base pour un jugement de valeur	Se réfère à des objectifs, à des normes ou à un programme rival	Se réfère aux différents systèmes de valeurs des personnes concernées
Préférences	Refuser de répondre plutôt que risquer de manquer d'objectivité	Sacrifier un peu la précision des mesures pour être plus utile
Précautions prises pour réduire les biais	Utiliser des procédures objectives et ne pas s'impliquer	Répéter les observations et chercher confirmation des conclusions

Critique des méthodes naturalistes

- Elles se contentent de restituer aux participants les informations qu'ils ont données eux-mêmes.
- Elles restent au niveau des impressions subjectives alors qu'on aimerait les dépasser.
- Elles sont difficilement planifiables.
- Elles peuvent prendre beaucoup de temps.

Épistémologie intersubjectiviste

La démarche anthropologique, illustrée par les modèles d'évaluation précédents, permet de connaître comment d'autres sujets vivent leur expérience, mais débouche difficilement sur une vision commune du curriculum à apprécier.

La vie en société, dans les petits groupes surtout, apprend cependant à forger des représentations collectives cohérentes, grâce à la confrontation des points de vue individuels. Sans atteindre l'objectivité, on peut au moins y tendre à travers l'intersubjectivité, en dégagant les traits communs aux diverses expériences

personnelles. C'est la stratégie des deux modèles suivants, qui visent chacun à leur manière à dépasser la subjectivité grâce au débat sur la base d'arguments étayés.

Utilités particulières

Les utilités restent particulières, comme les représentations initiales, mais elles aussi peuvent être confrontées, à la poursuite d'une politique consensuelle.

Orientation théorique: vers la connaissance

Évaluation judiciaire (non comparative)

Thomas Owens a proposé en 1973 que les aspects importants d'un curriculum à évaluer soient présentés par deux équipes, l'une chargée de présenter les aspects en faveur de son adoption, l'autre soutenant la thèse inverse. Ainsi tous les problèmes importants pourraient être abordés et discutés à fond.

La procédure utilisée est celle d'un procès judiciaire, qui permet d'éviter une approche unilatérale. Les problèmes délicats peuvent être creusés et discutés, par exemple la façon dont le programme a été compris et appliqué.

Il n'est pas question, en général, pour le juge d'aboutir à une décision pour ou contre le curriculum en question: l'essentiel est pour lui d'éclairer la situation, de savoir sur quoi les deux parties sont d'accord, comme les domaines où les avis divergent (et jusqu'à quel point).

Si c'est le décideur intéressé par le curriculum qui joue le rôle de juge, ce modèle se rapproche davantage des aides à la décision et doit être classé dans la catégorie suivante.

Orientation pratique: vers la décision

Évaluation négociée (non comparative)

Jacques Ardoino et Guy Berger sont les principaux défenseurs de ce modèle. Ils demandent aux délégués des groupes concernés par le fonctionnement d'un programme d'enseignement, de confronter leurs perceptions, puis de négocier des critères d'évaluation communs, de manière à arriver à des conclusions acceptables par tous, dans un cadre de référence élargi.

Le but n'est pas de faire la moyenne d'interprétations divergentes, mais plutôt de trouver un sens à ces oppositions, en dégagant les référentiels qui les expliquent. Des représentations contradictoires peuvent éclairer des aspects complémentaires, qu'il est utile de faire apparaître. La participation des évalués à leur évaluation est ainsi une condition de validité des résultats.

Par ailleurs, si la vérité n'est pas à négocier, la recherche de solutions efficaces peut l'être. Prendre en compte les valeurs des intéressés oblige à ouvrir un débat avec eux à quatre niveaux: définition du problème, recueil des informations, interprétation, plan de correction.

Les évalués, participant à la recherche, sont finalement les meilleurs évaluateurs, à condition que tous les points de vue soient pris en compte.

Discussion de cette typologie

Une façon de simplifier

Marc-André Nadeau présente d'autres modèles et la littérature scientifique en comporte encore bien plus. La typologie donne une base pour en choisir un seul par type.

Une proposition parmi d'autres

Une douzaine de typologies analogues ont déjà été publiées, visant aussi à organiser les innombrables méthodes proposées pour conduire des évaluations de programmes ou de curriculums (Voir Wang et Walberg. 1983). Toutes mettent en évidence certaines similarités.

Classement discutable

La décision de ranger tel ou tel modèle dans chaque case de la typologie comporte beaucoup d'arbitraire. Par certains traits, un modèle pourrait se rattacher à une autre case.

Son caractère plus structuré

L'intérêt de cette typologie est cependant de prendre en compte 4 dimensions de classement à la fois, et chacune de façon exhaustive. A la limite, les dix types qui en découlent définissent dix modèles, qui peuvent remplacer, en les résumant, tous les modèles existants.

Sa référence philosophique

On pourrait élaborer davantage les implications des dimensions proposées. Par exemple, Ardoino considère comme des formes de "contrôle" toutes les évaluations effectuées sur la base d'une épistémologie objectiviste, alors qu'il réserve le terme "évaluation" aux travaux qui se situent dans le cadre des deux autres épistémologies.

Ardoino et Berger développent cette distinction fondamentale dans l'introduction de "Une évaluation en actes" (1985), p. 9 à 21 (texte à lire).

Sa référence sociologique

Jean-Marie Barbier (dans "L'évaluation en formation" p. 250-257) rattache les modes d'évaluation correspondant aux trois épistémologies, à trois formes de rapport de pouvoir pouvant exister dans les organisations. Il cite lui-même Maurice Jeannet: "Le processus d'évaluation est intimement lié aux structures formelles et informelles caractérisant le milieu concret dans laquelle il se déroule. Plus exactement, il exprime ces structures et se trouve déterminé par elles. (...) La forme des relations de pouvoir, notamment, influence directement la manière dont l'évaluation est gérée." Barbier continue (p. 249): "Cette fonction de révélateur s'exerce notamment à deux occasions:

- A l'occasion de la définition des objectifs et des projets. Celle-ci ne peut se faire sans que ceux qui ont précisément le pouvoir de se prononcer sur ces projets ne soient obligés de le faire. L'évaluation peut dès lors révéler les manipulations, pressions, enjeux présents dans la situation et elle risque de faire apparaître au grand jour des tensions latentes.

- A l'occasion de l'appropriation des résultats. La connaissance des destinataires effectifs et des titres auxquels ils sont concernés permet en effet, (...) de lever d'éventuelles ambiguïtés sur l'exercice du pouvoir dans les actions."

le tableau 15 de Jean-Marie Barbier "Modes ou manières sociales d'évaluer les actions" recoupe particulièrement bien la typologie proposée (voir photocopie).

Son utilité pour choisir une méthode d'évaluation

La citation de Barbier montre bien que le choix du mode d'évaluation doit se faire en fonction des caractéristiques de la situation. Les exigences de l'instance qui demande et finance l'évaluation ne peuvent être négligées: on ne peut imposer une évaluation participative à un milieu social qui fonctionne selon une hiérarchie militaire. De même, la prise en compte, ou non, des utilités des divers partenaires dépend du contexte et de l'objet à évaluer. L'orientation peut aussi être plus ou moins théorique ou pratique, selon l'urgence des décisions à prendre.

L'intérêt de la typologie proposée est donc d'explicitier les choix fondamentaux à effectuer et, en les mettant en rapport avec les demandes du contexte particulier, de déterminer le modèle le plus approprié à la situation.

APPROCHE DESCRIPTIVE

Dans la réalité, les recherches évaluatives ne suivent pas toujours les modèles qui viennent d'être présentés. Certaines démarches constituent en fait des pseudo-évaluations.

Une des critiques faites à l'ensemble des modèles d'évaluation est qu'ils sont de toute façon peu utiles, parce que les décisions concrètes sont prises indépendamment des évaluations. Il vaut donc la peine d'aller voir comment se prennent les décisions.

Pseudo-évaluations

Recherches de légitimation

A côté des recherches qui préparent ou accompagnent des innovations, il peut exister des recherches qui visent à conforter la situation actuelle, et par là les structures de pouvoir en place et leur politique. L'établissement de statistiques sur le nombre de diplômes accordés dans le pays peut être de ce type. Les responsables utilisent les informations qu'ils tirent de ces recherches comme un instrument de pouvoir: ils peuvent en tirer parti dans leurs décisions; ils peuvent en diffuser certains éléments, dans le cas où ils leur sont favorables.

Mais les syndicats et les autres groupes de pression peuvent jouer le même jeu symétriquement: ils peuvent bloquer les recherches qui menacent de dévoiler des faits utilisables contre leurs intérêts. Dans un cas comme dans l'autre la recherche vise à fournir des armes qu'on garde en réserve parce qu'elles pourraient être utiles en cas de conflit.

Recherches de propagande

D'autres pseudo-recherches visent à fournir des éléments utiles pour une politique de relations publiques. Il s'agit de créer une image positive de l'institution ou du programme de formation, et l'étude recherchera les éléments qui rencontrent la faveur du public visé. Le directeur d'un programme demandera, par exemple, d'interroger les élèves sur les innovations dont il sait d'avance qu'elles sont bien accueillies.

Recherches fragmentaires

Certaines recherches comparatives classiques suivent un plan rigoureux, mais en se limitant à la mesure de quelques performances étroites, elles risquent de donner une image déformée de la valeur réelle des programmes comparés.

Certaines évaluations sommatives établissent un bilan correct en fin d'étude, mais les mesures obtenues ne renseignent nullement sur les directions d'amélioration possibles.

Evaluations diffuses*Modes de décision institutionnels*

Les décideurs semblent appliquer des procédures diamétralement opposées à celles que proposent les modèles prescriptifs:

- les buts, au lieu d'être précisés au départ, restent vagues, (parce que cela évite de créer des dissensions et que les décideurs ont ainsi une plus grande marge de manoeuvre).
- les moyens sont motivants par eux-mêmes et sont souvent choisis avant les buts, influençant donc le choix de ces derniers. (On veut des ordinateurs et l'on cherche ce que l'on pourrait faire avec eux en classe.)
- les changements se font par petites étapes, de sorte que le passé détermine les décisions plus que l'avenir souhaité: "on avance en regardant dans le rétroviseur."
- Les nouveaux principes donnent lieu à tous les ajustements "réducteurs" que l'autorité tolère. Ainsi, l'innovation est toujours "récupérée" pour que l'ancien fonctionnement puisse continuer (la sélection, par exemple).
- les contingences locales ont plus de poids que les structures communes à plusieurs sites, ce qui rend les généralisations relatives à un programme très incertaines.
- les décisions se font pas à pas et l'on refuse de prendre des décisions de principe à long terme (qui empêcheraient de revenir en arrière).
- Les objectifs changent donc du début à la fin d'une expérience et les points de repères font défaut pour apprécier les résultats.
- Les décisions sont prises progressivement par une évolution des attitudes dans le groupe, avant même le moment où elles sont sensées être prises, et souvent par d'autres personnes que les responsables en titre.

Rôles des évaluateurs

Il n'est pas étonnant dans ces conditions que les évaluateurs soient perçus comme des "extra-terrestres" par les enseignants, (qui tendent

à les rejeter, comme extérieurs au système). Les évaluateurs doivent souvent négocier leur participation.

Ils peuvent même être les pions que les groupes de pression essaient de manipuler pour qu'ils servent leurs intérêts.

Le plus qui puisse être fait du point de vue scientifique se limite souvent à une étude de cas descriptive, sans généralisation possible à d'autres situations. Le regroupement d'expériences aux déroulements non comparables n'aurait aucun sens, car "l'Histoire ne se répète pas."

Le contact des milieux de l'enseignement avec les spécialistes des sciences de l'éducation crée cependant une acculturation progressive et une lente diffusion des concepts de ces derniers dans le tissu social.

Intégration possible des approches prescriptives et descriptives

Peut-on sauver le modèle rationnel dans ce monde irrationnel?

Développement de logiques particulières

La meilleure façon de le faire serait de découvrir la logique sous-jacente au fonctionnement réel des institutions et de trouver des modes d'organisation qui les utilisent. Par exemple, les assurances répondent à la conduite "irrationnelle" des humains qui préfèrent des petites pertes certaines (les primes à payer) à une grosse perte improbable (l'incendie).

Utilité de guides pour promouvoir la rationalité

Dans l'attente de tels développements, les évaluateurs peuvent accorder une attention spéciale aux problèmes de communication avec les diverses "audiences" de leur étude. Les modèles naturalistes semblent préférables de ce point de vue.

Les modèles qui stimulent la participation des intéressés ont aussi plus de chances d'arriver à des conclusions pertinentes et à des décisions utiles.

Un modèle de décision rationnel peut servir de guide, même si l'on sait que d'autres facteurs entrent en jeu dans les décisions de groupes.

Vue d'ensemble sur les méthodes d'évaluation de curriculums

Points d'accord

En dépit des larges divergences d'opinion évoquées ci-dessus, les théoriciens de l'évaluation de curriculums s'entendent sur les huit conclusions ci-dessous, tirées de l'Encyclopedia of Educational Evaluation, p. 136:

- 1) La raison d'être de l'évaluation d'un programme éducatif est de fournir des informations qui permettent de prendre des décisions à son sujet. Par conséquent, une étude d'évaluation doit être planifiée en fonction des options entre lesquelles il faudra choisir.
- 2) Les résultats de l'évaluation devront pouvoir servir à améliorer le curriculum, et non seulement à l'accepter ou à le rejeter.
- 3) L'information devra être apportée assez tôt pour pouvoir fonder ces décisions.

- 4) L'évaluation est un jugement porté sur des résultats: les mesures ne font que préparer l'évaluation.
- 5) L'évaluation doit prendre en compte les objectifs à long terme autant que les objectifs à court terme, les effets imprévus aussi bien que les résultats attendus.
- 6) Les causes possibles autres que le programme pouvant avoir influencé les résultats doivent être prises en compte, (comme la motivation du groupe expérimental ou du groupe contrôle).
- 7) Les effets sont presque fatalement multiples, et les réactions des sujets presque fatalement différentes, ce qui implique des analyses d'interactions multiples également.
- 8) Les informations recueillies doivent satisfaire des critères minimaux de fidélité, de validité, d'utilité, et respecter des principes éthiques.

Points de divergence

Au-delà de ces principes, les démarches des évaluateurs peuvent varier sur les six dimensions suivantes, en fonction des aspects particuliers du problème posé:

- 1) Finalité sommative ou formative
- 2) Approche comparative ou non comparative
- 3) Evalueur intérieur ou extérieur à l'organisation
- 4) Etude de cas localisée ou recherche généralisable
- 5) Méthodologie quantitative ou qualitative
- 6) Plan prédéfini ou approche ouverte

LE RÔLE SOCIAL EFFECTIF DE L'ÉVALUATION DE CURRICULUM

Impact limité des évaluations effectuées

Complexité de la prise de décisions

Il va de soi que toute décision de type social fait intervenir beaucoup de facteurs souvent contradictoires, et que la qualité du curriculum ne peut pas être le seul aspect à considérer.

En particulier, il faut tenir compte de l'inertie des institutions: il faut beaucoup de preuves répétées pour les mettre en mouvement et de nombreux rapports négatifs pour les arrêter. On ne peut stopper un programme en cours d'exécution.

Revenir en arrière est impossible pour un décideur politique, car il perdrait la face; il faut que ce soit un parti opposé qui le fasse. Comme la Suisse connaît peu d'alternance, c'est en général en lançant un nouveau programme que l'on peut corriger l'ancien.

Prise en compte insuffisante des conclusions des évaluations

On a remarqué, par exemple, que les décisions du gouvernement américain ne semblaient pas avoir tenu compte des recherches faites pour évaluer les grands programmes télévisuels Head Start ou Sesame Street, tout au moins avant une dizaine d'années.

Ce désintérêt est encore plus manifeste chez les gouvernants européens. Michael Huberman, demandant aux Chefs de Départements suisses quel était le rôle de la recherche pédagogique pour eux, s'est fait répondre en privé qu'ils n'en tenaient pas compte...

Compétition d'autres motivations

Des curriculums peuvent être choisis ou rejetés en raison de l'implication personnelle des responsables à leur sujet (projets de carrière, par ex.), ou parce qu'ils contribuent à étendre l'influence d'une organisation, ou pour des raisons économiques ou idéologiques.

Insuffisance des rédactions de rapport de type psychométrique

Une des raisons du manque d'impact peut provenir de la maladresse des rapports qui adoptent un style de chercheurs.

Ni les enseignants, ni le public ne maîtrisent par exemple les concepts de centile, ou d'équivalent-degré, n'aperçoivent l'importance du choix de l'étalonnage, n'imaginent le caractère très partiel des connaissances évaluées par rapport à tout ce que connaissent les élèves. D'où souvent de graves erreurs d'interprétation. Les rapports ne sont lus que s'ils sont compréhensibles et ils ne sont appliqués correctement que s'ils transmettent clairement leurs conclusions.

Inadéquation du modèle décisionnel

Plus fondamentalement, il faut relativiser l'approche logique selon laquelle un décideur compare l'utilité des choix possibles et choisit le meilleur.

Multiplicité des buts

Jamais des buts ne sont acceptés unanimement. Ceux qui sont choisis le sont parce qu'ils répondent partiellement aux vœux de divers groupes, qui en attendent en fait des résultats contradictoires. Ex: la démocratisation des études à Genève, dont les classes privilégiées ont profité plus que les autres. - C'est pourquoi les buts sont exprimés en termes vagues pour couvrir ces ambiguïtés voulues.

Multiplicité des effets

Même les meilleurs plans expérimentaux ne peuvent dissocier l'effet de la conjoncture économique ou politique, des conflits sociaux ou de personnes, etc. Il est donc difficile de conclure de façon assurée pour le passé et de prédire les résultats futurs.

Multiplicité des préoccupations

Le temps nécessaire à une étude fait changer la problématique prioritaire et les résultats n'intéressent plus les décideurs quand ils sont publiés, parce qu'ils s'intéresseraient désormais à d'autres effets: Ex. La mathématique "moderne" dont l'évaluation a paru quand les politiciens s'intéressaient au français.

Multiplicité des conditions

Une organisation continue sur sa lancée tant que des pressions extérieures (campagne de presse, débat au Grand Conseil,...) ne la met pas en cause. Devant ces réclamations, elle lance un programme, par ex. d'évaluation. Six mois plus tard le programme commence, mais l'opinion publique se préoccupe d'autre chose déjà.

L'évaluation peut être faite de bonne foi, mais elle n'aura d'effet que si se trouvent réunies une série de conditions: des résultats nets, un

appui de l'opinion publique et des milieux politiques majoritaires, et une attitude réceptive de l'organisation.

Si une de ces conditions n'est pas présente, l'évaluation sera simplement classée.

Autres fonctions des évaluations pour l'organisation

Gestion des conflits

Annoncer une évaluation, pour l'instance au pouvoir, c'est signaler que la décision est encore sujette à négociation et à compromis.

Accepter une évaluation, pour les opposants, c'est dire que tout n'est pas nécessairement rejeté dans l'innovation.

Pour les deux parties, c'est se donner du temps et laisser se calmer les esprits. Le problème sera vite oublié, même par les plus fanatiques...

Réduction du laisser-aller

Savoir que le programme va être évalué oblige les participants au programme à surveiller de plus près son fonctionnement. La discussion de ses objectifs conduit à travailler de façon plus cohérente.

On a donc toute raison de penser que les résultats seront meilleurs. De plus, les participants seront plus motivés et donc plus satisfaits de leur travail.

Rituel social

Le modèle rationnel lié à l'évaluation donne une image rassurante des instances de décision. Le public devient moins anxieux. La société est plus unie derrière l'autorité.

Importance des évaluations comme source d'information générale

Les évaluations finissent par être entendues

Des recherches sur la diffusion et l'utilisation des savoirs nouveaux montrent que l'information diffuse lentement dans la société. Les résultats des évaluations peuvent changer graduellement la façon dont les problèmes sont envisagés dans les milieux où se prennent les décisions. Les gens deviennent plus "éclairés".

Rôle des sciences sociales

C'est même le mode particulier d'application des sciences sociales. Elles ne donnent pas de techniques directement applicables, chaque situation étant différente des autres. Par contre, elles contribuent à changer les représentations dans la société, et de ce fait elles amènent les gens à se comporter différemment. Elles ont ainsi un effet très marqué, mais indirect et diffus.

Bibliographie

- Barbier, J.-M. (1985). L'évaluation en formation. Paris: P.U.F.
- Nadeau, M.-A. (1988). L'évaluation de programme: théorie et pratique. Québec: Les Presses de l'Université Laval.
- Rossi, P. H., Freeman, H. E., & Wright, S. R. (1979). Evaluation: a systematic approach. Beverly Hills: Sage Publications.
- Stufflebeam, D. L., & Shinkfield, A. J. (1985). Systematic evaluation: a self-instructional guide to theory and practice. Boston: Kluwer-Nijhoff Publishing.
- Walberg, H. J., & Haertel, G. D. (1990). The international encyclopedia of educational evaluation. Oxford: Pergamon Press.

TABLEAU 15. — Modes ou manières sociales d'évaluer les actions

<i>Exemples de démarches et de méthodes d'évaluation</i>	Evaluations : - « prédéterminées », - de type instrumentaliste, techniciste ou formalisé, - de type fonctionnel, - à visée objectivante.	Evaluations : - centrées sur le groupe, - de type relationnel ou personnel, - de type institutionnel, - « réactionnelles », - de type « transaction-observation », - de type régulation.	Evaluations : - « à objectifs multiples », - « concurrentes », - « holistiques », - de type interactif.
<i>Rapports sociaux dans lesquels s'effectue l'évaluation</i>	Rôle dominant des responsables de l'action et de leur pouvoir de tutelle sans pouvoir direct des participants en tant que tels.	Rôle privilégié accordé au contraire aux participants à l'action.	Chaque type d'acteur impliqué dans l'action, et notamment les participants, joue un rôle de partie prenante parmi d'autres dans l'évaluation.
<i>Incidences sur le fonctionnement de l'évaluation :</i> <i>Au niveau du choix des objectifs et des indicateurs</i>	1 / Détermination <i>a priori</i> et stabilité relative. 2 / Privilège sensible accordé à la progression et aux résultats de l'action. 3 / « Directivité ».	1 / Indétermination au départ explicitation en cours d'évaluation. Possibilité de modification. 2 / Privilège accordé aux conditions de fonctionnement de l'action, aux processus. 3 / Libération de la parole des individus et des groupes.	1 / Développement parallèle de plusieurs processus de choix correspondant à chacun des types d'acteurs. 2 / Choix hétérogènes et morcelés, mais présentant une grande diversité. 3 / Expression autonome de chaque type d'acteur en tant que tel.
<i>Au niveau méthodologique de la production d'informations et de la mise au point de critères</i>	1 / Caractère externe et formalisé des outils. Méthodes objectivantes. 2 / Privilège relatif donné à l'observation et à la provocation de conduites (tests). Modèle expérimental. 3 / Perspectives de quantification et de mesure. Méthodes quantitatives.	1 / Outils susceptibles de favoriser l'expression notamment orale des aspirations individuelles ou collectives. 2 / Mise en œuvre potentielle de ces outils tout au long de l'action. 3 / Perspectives d'implication.	1 / Grande diversité des outils privilège relatif au travail de groupe. 2 / Distance sociale minimum entre évaluation et action. 3 / Caractère négocié de ces outils et d'une manière plus générale des conditions de l'évaluation.
<i>Au niveau du jugement de valeur proprement dit</i>	1 / Produit plutôt « en final » de l'action. 2 / Porté plutôt par les responsables de l'action. 3 / A trait surtout aux points et points faibles des résultats de l'action.	1 / Acte permanent. 2 / Porté essentiellement par les participants. 3 / Variable dans son contenu.	1 / Coexistence de jugements différents portés par les différents types d'acteurs. 2 / Ces jugements ont trait à des aspects très divers des actions. 3 / Ces jugements ne se présentent pas comme des jugements scientifiques, mais comme des jugements sociaux.
<i>Mode d'analyse des besoins correspondant</i>	Détermination des objectifs à partir des exigences de fonctionnement et de développement des organisations et systèmes.	Détermination des objectifs à partir de l'expression des aspirations, désirs, attentes des individus et des groupes.	Détermination des objectifs à partir de la définition des intérêts en situation.
<i>Mode de conduite des actions correspondant</i>	La population cible de l'action n'est pas directement associée en tant que telle à la conduite de l'action.	La population cible de l'action se voit octroyer un rôle apparemment privilégié dans la conduite de cette action.	La population cible de l'action occupe une position de partenaire dans la conduite de l'action.

TABLEAU 1.**Analyse des types d'études centrées sur les questions (Quasi-évaluation)**

<i>Approches</i>	<i>Orientation vers un questionnement spécifique (Quasi-évaluation)</i>				
<i>Définitions</i>	<i>Études qui abordent des questions précises et dont les réponses peuvent ou non établir la valeur (bonne ou mauvaise) d'un objet</i>				
<i>Types d'études</i>	<i>Études centrées sur les objectifs</i>	<i>Études pour rendre des comptes</i>	<i>Études avec recherches expérimentales</i>	<i>Programmes de passation de tests</i>	<i>Systèmes intégrés de gestion</i>
<i>Préoccupations</i>	Objectifs	Responsabilités de personnel ou d'établissement	Énoncés des problèmes, hypothèses et questions	Domaines du curriculum, tests publiés et groupes spécifiques de référence	Objectifs du programme, activités et événements
<i>Buts</i>	Rattacher les résultats aux objectifs	Fournir aux électeurs/mandants une comptabilité exacte des résultats	Déterminer la relation causale entre des variables indépendantes et dépendantes déterminées	Comparer les performances individuelles des élèves et celles des groupes d'élèves en vue d'établir des normes	Fournir continuellement l'information nécessaire pour financer, diriger et, contrôler les programmes
<i>Sources des questions</i>	Réalisateurs et directeurs de programmes	Électeurs/mandants	Chercheurs et réalisateurs	Éditeurs de tests et commissions de sélection de tests	Personnes dirigeantes
<i>Questions principales</i>	Quels élèves atteignent quels objectifs ?	Est-ce que les personnes et les organismes auxquels on confère des responsabilités accomplissent tout ce qu'ils devraient accomplir ?	Quels sont les effets d'une intervention donnée sur des variables déterminées de résultats ?	Les performances individuelles des élèves se situent-elles au même niveau ou au-dessus de la performance moyenne du groupe de référence ?	Les activités du programme sont-elles mises en oeuvre dans les délais fixés, à un coût raisonnable et avec les résultats attendus ?
<i>Méthodes typiques</i>	Analyse des résultats obtenus relatifs à des objectifs déterminés	Procédures d'audit et programmes de passation de tests sur mandat	Plan expérimental et quasi-expérimental	Choix, passation, notation de tests standardisés et présentation des résultats	Système d'analyse PERT (Program Evaluation and Review Technique = méthode des potentiels), CPM, PPBS, systèmes informatisés d'information et analyse des coûts
<i>Pionniers</i>	Tyler	Lessinger	Lindquist, Campbell and Stanley, and Cronbach	Lindquist	Cook, Kerr
<i>Développement assuré par</i>	Bloom, Hammond, Metfessel & Michael, Popham and Provus	Stenner and Webster, Kearney	Suchman, Wiley, Glass	Flanagan, Lord and Novick, Hyronymous, Thorndike, and many more	Kaufman

[Référence : *Stufflebeam, D. & Webster, W. (1983).*]

TABLEAU 1.**Analyse des types d'études centrées sur les questions (Quasi-évaluation)**

<i>Approches</i>	<i>Orientation vers un questionnement spécifique (Quasi-évaluation)</i>				
<i>Définitions</i>	<i>Études qui abordent des questions précises et dont les réponses peuvent ou non établir la valeur (bonne ou mauvaise) d'un objet</i>				
<i>Types d'études</i>	<i>Études centrées sur les objectifs</i>	<i>Études pour rendre des comptes</i>	<i>Études avec recherches expérimentales</i>	<i>Programmes de passation de tests</i>	<i>Systèmes intégrés de gestion</i>
Préoccupations	Objectifs	Responsabilités de personnel ou d'établissement	Énoncés des problèmes, hypothèses et questions	Domaines du curriculum, tests publiés et groupes spécifiques de référence	Objectifs du programme, activités et événements
Buts	Rattacher les résultats aux objectifs	Fournir aux électeurs/mandants une comptabilité exacte des résultats	Déterminer la relation causale entre des variables indépendantes et dépendantes déterminées	Comparer les performances individuelles des élèves et celles des groupes d'élèves en vue d'établir des normes	Fournir continuellement l'information nécessaire pour financer, diriger et, contrôler les programmes
Sources des questions	Réalisateurs et directeurs de programmes	Électeurs/mandants	Chercheurs et réalisateurs	Éditeurs de tests et commissions de sélection de tests	Personnes dirigeantes
Questions principales	Quels élèves atteignent quels objectifs ?	Est-ce que les personnes et les organismes auxquels on confère des responsabilités accomplissent tout ce qu'ils devraient accomplir ?	Quels sont les effets d'une intervention donnée sur des variables déterminées de résultats ?	Les performances individuelles des élèves se situent-elles au même niveau ou au-dessus de la performance moyenne du groupe de référence ?	Les activités du programme sont-elles mises en oeuvre dans les délais fixés, à un coût raisonnable et avec les résultats attendus ?
Méthodes typiques	Analyse des résultats obtenus relatifs à des objectifs déterminés	Procédures d'audit et programmes de passation de tests sur mandat	Plan expérimental et quasi-expérimental	Choix, passation, notation de tests standardisés et présentation des résultats	Système d'analyse PERT (Program Evaluation and Review Technique = méthode des potentiels), CPM, PPBS, systèmes informatisés d'information et analyse des coûts
Pionniers	Tyler	Lessinger	Lindquist, Campbell and Stanley, and Cronbach	Lindquist	Cook, Kerr
Développement assuré par	Bloom, Hammond, Metfessel & Michael, Popham and Provus	Stenner and Webster, Kearney	Suchman, Wiley, Glass	Flanagan, Lord and Novick, Hyronymous, Thorndike, and many more	Kaufman

[Référence : *Stufflebeam, D. & Webster, W. (1983).*]

TABLEAU 1.**Analyse des types d'études centrées sur les questions (Quasi-évaluation)**

<i>Approches</i>	<i>Orientation vers un questionnement spécifique (Quasi-évaluation)</i>				
<i>Définitions</i>	<i>Études qui abordent des questions précises et dont les réponses peuvent ou non établir la valeur (bonne ou mauvaise) d'un objet</i>				
<i>Types d'études</i>	<i>Études centrées sur les objectifs</i>	<i>Études pour rendre des comptes</i>	<i>Études avec recherches expérimentales</i>	<i>Programmes de passation de tests</i>	<i>Systèmes intégrés de gestion</i>
Préoccupations	Objectifs	Responsabilités de personnel ou d'établissement	Énoncés des problèmes, hypothèses et questions	Domaines du curriculum, tests publiés et groupes spécifiques de référence	Objectifs du programme, activités et événements
Buts	Rattacher les résultats aux objectifs	Fournir aux électeurs/mandants une comptabilité exacte des résultats	Déterminer la relation causale entre des variables indépendantes et dépendantes déterminées	Comparer les performances individuelles des élèves et celles des groupes d'élèves en vue d'établir des normes	Fournir continuellement l'information nécessaire pour financer, diriger et, contrôler les programmes
Sources des questions	Réalisateurs et directeurs de programmes	Électeurs/mandants	Chercheurs et réalisateurs	Éditeurs de tests et commissions de sélection de tests	Personnes dirigeantes
Questions principales	Quels élèves atteignent quels objectifs ?	Est-ce que les personnes et les organismes auxquels on confère des responsabilités accomplissent tout ce qu'ils devraient accomplir ?	Quels sont les effets d'une intervention donnée sur des variables déterminées de résultats ?	Les performances individuelles des élèves se situent-elles au même niveau ou au-dessus de la performance moyenne du groupe de référence ?	Les activités du programme sont-elles mises en oeuvre dans les délais fixés, à un coût raisonnable et avec les résultats attendus ?
Méthodes typiques	Analyse des résultats obtenus relatifs à des objectifs déterminés	Procédures d'audit et programmes de passation de tests sur mandat	Plan expérimental et quasi-expérimental	Choix, passation, notation de tests standardisés et présentation des résultats	Système d'analyse PERT (Program Evaluation and Review Technique = méthode des potentiels), CPM, PPBS, systèmes informatisés d'information et analyse des coûts
Pionniers	Tyler	Lessinger	Lindquist, Campbell and Stanley, and Cronbach	Lindquist	Cook, Kerr
Développement assuré par	Bloom, Hammond, Metfessel & Michael, Popham and Provus	Stenner and Webster, Kearney	Suchman, Wiley, Glass	Flanagan, Lord and Novick, Hyronymous, Thorndike, and many more	Kaufman

[Référence : *Stufflebeam, D. & Webster, W. (1983).*]

CH. 5B. DISCUSSION DE LA RECHERCHE LECTURE

INTRODUCTION¹ :

PRÉSENTATION DES 3 THÈMES DE DISCUSSION :

Après avoir lu des ouvrages, il arrive parfois que l'on aimerait bien voir le ou les auteurs et leur poser quelques questions.

Aujourd'hui, nous avons la chance de pouvoir réaliser ce désir, car l'un des auteurs de la recherche effectuée sur les méthodes de lecture est totalement à notre disposition.

Les questions que je vais poser portent sur 3 thèmes :

- Le premier concerne la méthodologie de la recherche.
- Le deuxième, le contexte historique de la recherche.
- Et le troisième aborde l'utilité de la recherche de l'IRDP.

INDICATION SUR LA MANIÈRE DONT SERA MENÉE CETTE PARTIE DU COURS :

Je vous invite à profiter de l'occasion qui vous est présentée en complétant mon propre questionnaire.

En fait, je vous demande même de m'aider à poser des questions intéressantes et, pourquoi pas, à poser des questions sur les questions.

Au gré des réponses qui seront données, vous pouvez intervenir à tout moment.

CONTESTATION DE LA MÉTHODE EXPÉRIMENTALE :

La recherche-lecture de l'IRDP est une étude dans laquelle, par le moyen des corrélations, on compare deux ou plusieurs méthodes de lecture.

On se souvient que la méthodologie expérimentale classique est contestée par Lee Cronbach qui reproche à ce type d'évaluation le fait que les facteurs d'apprentissage ne sont finalement pas mieux compris.

Cronbach sera aussi l'un des premiers chercheurs à distinguer l'évaluation de la recherche.

¹ En style normal, les éléments de réponse, les explicitations supplémentaires et le discours prévu. En italique, les questions posées.

Soulignant également les insuffisances de la recherche expérimentale traditionnelle, les auteurs nous proposent une nouvelle méthodologie : l'approche systémique comparative.

En effet, parmi les principaux reproches faits à l'approche traditionnelle, on peut relever que variables dépendantes (apprentissage des élèves) et variables indépendantes (méthodes de lecture) posent de sérieuses difficultés, par exemple :

- Il n'est pas possible de neutraliser toutes les autres variables causales (milieu scolaire et social, niveau intellectuel, sexe, âge, etc.)
- Il n'est pas possible non plus de contrôler entièrement la variable indépendante : dans le réel, cette variable n'existe pas : elle est composée d'un grand nombre de sous-variables (par exemple, la personnalité de l'enseignante, le nombre d'années d'expérience professionnelle, la formation initiale, le perfectionnement ultérieur, la provenance sociale, etc.).
- La variable-critère (c'est-à-dire la réussite scolaire en lecture) pose également problème, car des méthodes différentes se donnent des objectifs différents à atteindre.
- Une dernière considération est l'impossibilité de généraliser les résultats à d'autres situations.

En résumé, la nouvelle approche remet donc en question d'une part la causalité (du type si...alors), à laquelle elle oppose la finalité, et d'autre part le réductionnisme à laquelle elle oppose la globalité.

I) LA MÉTHODOLOGIE DE LA RECHERCHE :

SUIVANT CETTE APPROCHE SYSTÉMIQUE COMPARATIVE :

Va-t-on vers une explication ou vers la prise de décision ? Vers le sens ou vers l'action ?

On va vers la décision. Il s'agit de donner des conseils. On ne peut pas atteindre la vérité. Cependant on peut mettre en évidence des probabilités : certaines méthodes sont le plus souvent associées à de meilleurs résultats et inversement. Ces effets ne sont pas dus au hasard.

On va vers l'action. Il s'agit de proposer aux autorités une action meilleure que l'autre. Il s'est avéré impossible d'analyser les effets de la méthode uniquement. En effet, les sources d'influence multiples ne permettent pas de déterminer une cause unique.

LES PRINCIPALES INSUFFISANCES :

EN CE QUI CONCERNE LA COMPARAISON DES MÉTHODES :

Tous les tests de signification ont porté sur les classes. La différence de moyenne entre les deux ensembles de classes comparés est généralement faible par rapport aux variations existant à l'intérieur des classes.

- *Quelles conséquences peut-on tirer de cette remarque en ce qui concerne la recherche comparative des méthodes de lecture ?*

Même si les différences entre les classes sont minimales, elles se sont révélées significatives du fait du grand nombre de classes évaluées (120).

Le fait de ne pas avoir atteint les facteurs vraiment importants et décisifs justifiait-il de ne rien dire ? Il faut distinguer les variables changeables (méthodes, activités) et non changeables (niveau socio-économique, personnalité et enthousiasme de l'enseignant, contexte) afin d'agir sur ce qui peut être changé.

- *Des enquêtes cliniques dans les classes n'auraient-elles pas alors été nécessaires ? (Ex. la prise en compte des éléments subjectifs de l'apprentissage, les réactions des participants à l'expérience, le surinvestissement des enseignantes, etc.)*

Certaines variables changeables n'ont peut-être pas été aperçues. L'enquête clinique aurait pu être plus efficace. Cependant, certaines précautions ont été prises, notamment la collaboration avec une inspectrice ayant une longue expérience de l'enseignement.

Aujourd'hui, on utiliserait de préférence la technique de l'entretien à celle du questionnaire pour identifier plus sûrement les opinions des enseignants. La méthode employée à l'époque prévoyait tout, y compris les conclusions, que les hypothèses soient confirmées ou infirmées.

Les hypothèses dépendent du type de recherche. Si on recherche des lois, des causes, etc., l'approche expérimentale doit être rigoureuse et les conclusions prévues. Si on veut explorer un domaine, l'approche doit être qualitative et non-expérimentale. L'intuition et les idées du chercheur restent ouvertes à d'autres variables.

- *Peut-on parler plutôt de recherche quasi-expérimentale, car la comparaison des deux groupes est critiquable puisque l'on n'a pas pu constituer au départ des groupes totalement équivalents ?*

L'approche expérimentale a été critiquée. Cependant la recherche lecture fait bien partie des approches comparatives.

- *Pourquoi s'est-on intéressé davantage aux options pédagogiques plutôt qu'aux méthodes elles-mêmes et quels ont été les limites du questionnaire utilisé ? (A-t-on pu définir strictement en quoi consistaient les méthodes comparées ? Sur quoi portait la différence d'un point de vue pédagogique ?). Problème des étiquettes : méthode énoncée et méthode pratiquée.*

Il y a eu une confusion des variables indépendantes (âge et expérience de l'enseignant, milieu rural ou urbain, méthode) qu'il a été impossible de dissocier en soi. Il a donc fallu faire une analyse plus fine. Associer une étiquette à une méthode n'implique pas que cette étiquette corresponde à la méthode utilisée en classe. En fait, les enseignantes n'appliquent pas vraiment une méthode; d'autres éléments sont souvent plus importants (p. ex. le style d'enseignement). Des résultats scientifiques pourraient être apportés en allant voir de plus près les pratiques réelles des enseignantes.

La question a été posée autrement. Plutôt que de proposer une nouvelle méthode, il s'agissait de définir quels sont les éléments importants qui favorisent l'apprentissage de la lecture.

- *Mais dans ce cas, où se situe la pertinence d'une approche dite comparative ? On pensait que les propriétés de la méthode de lecture dérivait plutôt des conditions dans lesquelles elle est mise en oeuvre. (= il y a ainsi reformulation de la question par les chercheurs). Le questionnaire permettait d'obtenir, non pas deux méthodes, mais un continuum de méthodes. Les résultats montrent alors que l'une n'est pas meilleure que l'autre, mais que des facteurs jouent un rôle ...etc. (on confirme l'hypothèse).*

- *Pourquoi parle-t-on d'approche comparative ? Était-on prisonnier d'un cadre implicite de référence ?*

(Note : Juger qu'une petite voiture est un avantage ou un inconvénient par rapport à une grande voiture dépend en premier lieu de ce que recherche le conducteur. On voit bien que le choix repose sur un système de valeurs attribué aux caractéristiques réelles du véhicule. Ce système de valeurs est également attribué en fonction de quelque chose !).

Il n'est pas juste de dire que la méthode comparative n'a pas été utilisée. En fait, plusieurs types de méthode ont été comparés, notamment globale, syllabique et "S'exprimer-Lire", mais par l'intermédiaire de leurs composantes méthodologiques. De l'effet de ces dernières, on pouvait remonter aux méthodes.

- *Dans une telle approche comparative, n'y a-t-il pas assimilation entre ce qui est mesuré (l'aptitude à lire) et le producteur supposé de cette aptitude (la méthode choisie) ?*

(Point non traité)

- *Qu'en est-il de la validité des tests de lecture utilisés ? (Ex. type papier/crayon)*

La réussite en lecture peut être biaisée par le type de test utilisé. Ainsi, l'aptitude à coder la réponse est parfois plus importante que la compréhension.

- *A-t-on pu, par ces tests, évaluer réellement ce que l'on déclarait évaluer ?*

Chaque fois qu'on opérationnalise une variable, il y a des pièges. La méthode d'analyse factorielle est la meilleure façon d'investiguer ce qu'on mesure (ex : les corrélations entre un test de maths et des épreuves de vocabulaire, calcul numérique, raisonnement logique permettent de découvrir les composantes spectrales et cognitives du test).

- *Enfin, l'arsenal des traitements statistiques n'avait-il pas une valeur plutôt symbolique ?*

Lors de la présentation des résultats, les autorités et les enseignants n'ont pas compris les données statistiques. Il y a effectivement des aspects affectifs derrière les aspects abstraits et cognitifs des statistiques:

- les chercheurs sont rassurés : ils ont quelque chose à présenter.
- ils sont mis en valeur : l'aspect scientifique donne du poids et du sérieux aux résultats.

- ils sont protégés : personne ne comprend donc aucune critique n'est possible.

- Dans ce type de recherche, quels sont les liens avec la théorie générale des systèmes ? *Dans quel sens le terme systémique est-il pris ?* (Système ouvert; système fermé; cybernétique).

L'histoire de la systémique remonte aux années 50. Les systèmes sont des parties qui fonctionnent ensemble et pas en addition. Cette conception est à rapprocher de la cybernétique (science des régulations). Elle était valable en biologie, économie et électronique.

Dans les années 60, on assiste au développement de l'analyse des systèmes. On l'utilise pour mettre au point de grandes organisations (radar pour aéroport). Le travail des ingénieurs était quantitatif (mesurer avec précision les marges de variation). Dans cette optique, les systèmes sont fermés.

La psychologie s'est inspirée de cette théorie pour proposer la théorie générale des systèmes. Le modèle des systèmes est un modèle métaphorique qui permet d'approcher qualitativement les interactions et les rétroactions. Le modèle développé dans les années 60 considérait que les systèmes étaient fermés (régularités, précision). Au contraire, les systèmes analysés en psychologie sont ouverts : il y a toujours de l'imprévu; l'objectif fixé au départ est souvent remis en cause (p. ex. vie de famille).

L'approche utilisée pour la recherche lecture est bien systémique, mais dans le sens de systèmes relativement fermés (contrôler ce qui se passe d'un point de vue quantitatif). Afin d'éviter des confusions, il aurait été préférable de parler plutôt d'analyse de pistes causales ou de corrélations (étude d'un certain nombre de variables en interaction).

- *Cette recherche ne tente-t-elle pas de réunir deux approches théoriques différentes, l'expérimentale et la systémique ?*

(Point non traité)

EN CE QUI CONCERNE LA PUBLICATION DES RÉSULTATS :

- *Quel a été l'impact d'un tel rapport scientifique ?*

L'impact direct a été nul, personne n'a compris.

PROBLÈMES D'INTERPRÉTATION :

- *Quels ont été les problèmes soulevés par un tel rapport du point de vue des autorités, des enseignants et des spécialistes (logopédistes, orthophonistes, etc.) ?*

Les autorités n'en ont pas tiré de conclusions. Les enseignants ne voyaient pas le lien entre les facteurs tirés, les méthodes et leur pratique journalière. Les spécialistes (orthophonistes, logopédistes) souhaitaient obtenir des recettes et des indications sur les causes des difficultés. Sur ce point les chercheurs n'ont rien apporté, en fait ce n'était pas leur but.

PERTINENCE :

- *L'évaluation et les résultats obtenus ont-ils répondu à la demande énoncée par les décideurs et celle énoncée ensuite par les chercheurs (Légitimité de la reformulation de la question ?)?*

Le problème a été reformulé par les chercheurs afin d'augmenter la valeur des résultats. Au départ, la demande réelle des autorités était plus complexe que celle formulée (trouver une méthode de lecture pour la Suisse romande). Des problèmes politiques sont apparus. Il était difficile de mettre d'accord tous les cantons romands. La question implicite des autorités était de trouver un moyen de sortir de cette impasse. Les chercheurs ont donc préféré une approche multidimensionnelle. L'analyse des options pédagogiques et des pratiques des enseignantes devait permettre de formuler de nouvelles propositions. Cela évitait l'écueil de proposer une méthode déjà existante qui aurait inévitablement contrarié un des cantons romands. L'option choisie était scientifiquement défendable. La recherche a permis de répondre à la question des chercheurs, pas à la question explicite des autorités, mais bien à leur question implicite.

EN CE QUI CONCERNE L'UTILISATION ULTÉRIEURE DES RÉSULTATS :

- *Le temps mis pour exécuter cette recherche a-t-il profité à son utilisation ultérieure ?*

Il est difficile de répondre à cette question sans faire allusion au contexte politique. La recherche s'est déroulée au début de la coordination scolaire romande qui était alors engagée dans des conflits à propos des maths modernes. La rénovation du curriculum de français ne pouvait démarrer à cause des conflits sur les méthodes de lecture. La méthode globale utilisée à Genève dix ans auparavant avait créé un problème politique majeur. Les parents et un parti politique engagé

dans la campagne pour les élections étaient contre la méthode globale. Ce parti a remporté les élections. La méthode globale a donc été écartée et remplacée par une méthode syllabique. Ce conflit a créé un traumatisme important dans toute la population (enseignants, parents, politiciens) en même temps qu'un obstacle à la coordination. Le canton de Genève ne voulait pas revivre la même situation dix ans après. Des inquiétudes se sont manifestées vis-à-vis de ce qui allait être proposé. Dans ce contexte, il était impensable de réintroduire la méthode globale.

Dans le canton de Neuchâtel, une nouvelle méthode de lecture ("S'exprimer-Lire") était utilisée dans les classes. Bien qu'elle ne puisse être qualifiée de méthode globale, il existait une certaine méfiance : le risque était grand que cette méthode soit étiquetée comme "globale". Les chercheurs ont donc dû trouver une évaluation qui permette de sortir de l'impasse. Cela explique aussi la durée de la recherche (3 ans). En fait, les autorités n'ont pas poussé les chercheurs à aller plus vite. Le problème de la lecture était ainsi "mis au frais" : "Les chercheurs s'en occupent, pendant ce temps, on fait des maths".

L'évaluation de la lecture constituait donc une forme de recherche alibi (autrement dit : "Comment noyer le poisson et s'en sortir dignement"). En prenant du temps, on permettait au conflit de s'estomper. Quand plus personne ne parlerait du problème, on pourrait plus facilement proposer une nouvelle méthode. En effet, quand le rapport est sorti, la sensibilisation était moins forte pour les méthodes de lecture; les esprits étaient surtout occupés par les maths modernes.

Pour ménager les sensibilités locales, les chercheurs ont pris soin de ne pas dire, dans tous les résultats, que la méthode de Neuchâtel était bonne (exacerbation de la compétition entre les cantons) ou pas bonne (insultant pour Neuchâtel).

Les chercheurs étaient donc confrontés à un problème de diplomatie. C'est pourquoi la question a été reformulée. On a cherché à voir les effets de différentes options et actions pédagogiques (prérequis, enrichissement, partir de l'oral, expérience de l'enseignant). La méthode "S'exprimer-Lire" a été analysée en éléments et non pas globalement.

- *La recherche a-t-elle empêché des changements ? (A-t-elle causé par exemple, la remise en question ou même une marche arrière par les politiques de la mise en place des innovations).*

Avec les résultats d'une évaluation deux cas peuvent se présenter : soit on fait marche arrière, soit au contraire on accélère la réalisation des nouvelles directives. Mais il est très difficile de remettre en question des décisions prises antérieurement (cf. recherches en psychologie sociale), même lorsque ces décisions s'avèrent erronées. Un homme politique ne veut pas donner l'impression qu'il s'est trompé. La seule possibilité pour lui est de dire qu'il "corrige le tir".

Il applique en fait le principe de l'évaluation formative et il demande aux chercheurs les éléments pour le faire. Le but de la recherche est de faciliter le repérage des problèmes.

L'opinion publique, au contraire, évalue en termes sommatifs : les gens sont "pour" ou "contre" une innovation. Les journaux ne se centrent même que sur les difficultés. Pour la mathématique par exemple, la "Feuille d'avis de Neuchâtel" titrait : "30% des enseignants sont contre les mathématiques modernes" (souvent, on peut mettre en lumière certains aspects et faire dire aux chiffres ce qu'on veut).

Les pédagogues ont du mal à lutter contre les a-priori négatif. Pour le français, on a été très près du blocage (à Neuchâtel, blocage de l'innovation pendant deux ans; pour les autres cantons, après des discussions dans tous les parlements, cela s'est joué à quelques voix près).

Les recherches risquent d'être tantôt des alibis pour gagner du temps, tantôt involontairement l'occasion de publier des résultats repris par la presse et utilisés au sens contraire de ce qui avait été prévu. Dans le code de déontologie américain, le chercheur a la responsabilité non seulement de dire des choses vraies, mais encore de s'assurer d'être compris par les journalistes, les parents et les enseignants. Cette responsabilité doit tenir compte des aspects de la réalité sociale, économique et politique du moment.

LES BASES THÉORIQUES DE LA "MÉTHODE DE LECTURE" CHOISIE
PAR MAÎTRISE DU FRANÇAIS :

Selon Martine Wirthner, dans "Savoir savant - Savoir d'expérience", Cahier du Groupe des Chercheurs Romands et publié par l'IRDP en août 1991, les principes et les conclusions de la recherche de l'IRDP sur la méthode "S'exprimer-Lire" auraient été repris par Maîtrise du Français :

- 1) La motivation joue un rôle très important dans l'apprentissage de la lecture;
- 2) La lecture doit être une activité finalisée, à envisager selon l'approche large (contextualisée);
- 3) Il est préférable de partir de ce que connaît l'enfant, à savoir l'oral, pour aller vers l'inconnu (ou le moins bien connu, l'écrit). Il ne s'agit pas cependant de considérer l'un comme plus important que l'autre.
- 4) L'enseignement doit être différencié;
- 5) L'enfant devrait bénéficier de deux ans (jusqu'à la fin de 2P) pour apprendre à lire.

- *Mais quelles sont en réalité les relations effectives de ces choix avec les résultats tirés de la recherche ?*

Il est justifié de reconnaître le rôle d'une recherche ancienne. Ici, l'auteure n'a pas eu tort de souligner des convergences qui mettaient en valeur l'Institut où elle travaille. Cela ne veut pas dire que sans la recherche lecture, Martine Wirthner n'aurait pas dit la même chose. Les idées développées étaient très générales et admises par tout le monde.

- *D'une manière plus générale, quels sont le rôle et la responsabilité des chercheurs dans les choix opérés ultérieurement ?*

(Point non traité)

2) LE CONTEXTE HISTORIQUE ET SON INFLUENCE :

- *Quels étaient le contexte politique et les véritables enjeux ?*
- *Quelle a été finalement l'utilité effective de la recherche d'évaluation de l'IRDP dans ce contexte politique particulier ?*
- *Par exemple, a-t-elle contribué à rehausser le statut de la Coordination ?*

(Points traités plus loin)

LE CONTEXTE SCIENTIFIQUE : LES QUERELLES D'ÉCOLE ?

- *En pédagogie : peut-on dire que les choix ultérieurs apparaissent comme la reprise, par les autorités, de la méthode Freinet (passage par l'encodage !) généralement perçue comme une propagande politique de gauche ?*

Les aspects mis en évidence ont bien une familiarité avec la méthode Freinet, dont on sait qu'elle est teintée politiquement. Mais dans les classes Freinet, il y a beaucoup de points positifs : pédagogie active, enfants heureux. L'idée méritait d'être exploitée. Les spécialistes de la lecture ont cherché à glisser la méthode Freinet dans l'officialité, en gardant ce qui était bon, mais en éliminant l'aspect politique.

- *En psychologie : vouloir faire découvrir le fonctionnement de la langue plutôt que de l'imposer, tout comme on se propose de faire découvrir les structures logico-mathématiques, est-il encore aujourd'hui aussi novateur et pertinent qu'il le fût ?*

Cela reste pertinent, même si ce n'est plus novateur ni audacieux.

- *Y a-t-il eu contamination des vues prédominantes de l'époque sur les méthodes d'enseignement ?*

Les pédagogues ont plutôt eu tendance à brandir Piaget pour faire passer les idées qui étaient les leurs de toute façon. Ils ont même parfois dénaturé l'apport de Piaget.

En sociologie : on se proposait de lutter contre le handicap scolaire des classes défavorisées en partant de leur langage.

- *N'a-t-on pas au contraire accentué ce handicap en supprimant les manuels de référence pour les familles ?*
- *En fait, les querelles étaient-elles davantage de l'ordre de l'idéologique ou de l'ordre du scientifique ?*

Il faut être attentif aux effets contraires possibles lorsqu'on fait une recherche. Pour la lecture, la nouvelle méthode n'est pas nécessairement défavorable aux classes défavorisées. Elle peut les aider sur certains points : signification de ce qui est fait en classe (partir de l'oral) et motivation. D'autres aspects par contre ont un effet plutôt négatif : le fait qu'il n'y ait pas de manuel pourrait empêcher les parents d'aider leurs enfants, mais en réalité, l'absence de manuel ne change rien à l'aide apportée par les parents des enfants de classes défavorisées. Par contre, il est vrai qu'on demande aux enfants beaucoup d'initiatives qui sont généralement moins faciles pour les enfants de milieu défavorisé. Ces derniers ne sont pas assez aidés par la nouvelle méthode qui fait appel à l'autonomie, la

participation, l'imagination, alors que leur éducation familiale valorise plus l'obéissance et la mémorisation.

Les résultats individuels, non publiés, ont montré qu'il existe une interaction aptitude/traitement. Les enfants les plus défavorisés (10% de la population) réussissent mieux avec la méthode syllabique, alors que tous les autres réussissent mieux avec une méthode à départ oral. La méthode romande n'en a pas tenu compte, puisque les résultats n'étaient à l'époque pas disponibles. Quoi qu'il en soit, aucune méthode n'est la meilleure pour toutes les situations.

3) DE L'UTILITÉ EFFECTIVE DE LA RECHERCHE D'ÉVALUATION DE L'IRDP.

- *Finally, la recherche a-t-elle été profitable sur le plan de la pratique pédagogique ? Par exemple, a-t-elle répondu aux attentes des enseignantes ?*
- *N'a-t-elle pas joué plutôt un rôle d'arbitre entre les spécialistes ?*
- *A-t-elle aussi renforcé l'idée d'une gestion scientifique de l'école ?*
- *A-t-elle surtout contribué à assurer la crédibilité des chercheurs de l'IRDP auprès des décideurs ? (les DIP)*
- *A-t-elle aussi contribué à légitimer les décisions prises par les décideurs auprès du public ?*

Elle a été utile à la coordination romande en évitant un conflit entre les différents cantons à propos des méthodes. En montrant que les méthodes n'avaient pas forcément beaucoup d'importance, mais que d'autres facteurs interviennent sur l'apprentissage de la lecture, elle a calmé les tensions.

A quoi sert la Recherche ? Une recherche toute seule ne sert à rien, mais s'il n'y a pas de recherche, les idées n'avancent pas. La recherche a un effet indirect par relais successifs. Elle apporte des idées nouvelles qui ne passent pas tout de suite dans la pratique, mais, par un effet de généralisation, elle "lance une mode". Les recherches font évoluer l'école.

EN CONCLUSION :

(Les questions suivantes étaient prévues, mais n'ont pas pu être discutées faute de temps. Il reste intéressant d'y réfléchir pour effectuer une critique de cette recherche.)

LA MÉTHODE SYSTÉMIQUE COMPARATIVE PARTICIPE D'UNE ÉPISTÉMOLOGIE OBJECTIVISTE :

- *Dans ce type d'approche, quels cadres implicites définissent les problèmes, la méthodologie et les solutions pour les chercheurs ?*
- *Qu'entend-on par "scientifique" ? Le scientifique se définit-il uniquement par une méthode hypothético-déductive ?*
- *Les préoccupations des autorités n'ont-elles pas reçu plus d'attention que les préoccupations des enseignantes et que le problème des méthodes lui-même ?*
- *A-t-on répondu à la demande ? A quel type de demande a-t-on véritablement répondu ?*
- *L'efficacité de la connaissance intuitive qui résulte de l'habitude du contact avec les enfants aurait-elle pu être exploitée par la méthodologie énoncée ? Sinon, par quelle méthodologie, par quels types de savoirs aurait-on pu le faire ?*
- *Par exemple, dans le cas de la recherche-lecture, a-t-on réussi à faire comprendre ce que révélait l'expérience des enseignantes ?*
- *Si cela avait été réalisé, quelle influence cela aurait-il eu sur les choix ultérieurs et sur la polémique politique et scientifique entre deux ou plusieurs méthodes de lecture ?*

CRITIQUE GÉNÉRALE DE CETTE RECHERCHE-ÉVALUATION :

- *Avant la recherche-lecture, s'est-on posé les questions suivantes :*
- *A-t-on explicité les questions, les attentes et les référentiels (cadre de référence implicite, système de valeurs) d'évaluation des divers partenaires ?*
- *L'orientation de la question est-elle théorique ou pratique ?*
- *Cherche-t-on une réponse universelle ou locale ?*
- *L'étude a-t-elle des implications pratiques ?*
- *L'étude a-t-elle des effets directs, indirects ?*
- *Le coût de la recherche est-il raisonnable ?*
- *Que n'a-t-on pas cherché à savoir ?*
- *A qui s'adresse-t-on réellement ? A qui la recherche doit-elle s'adresser ?*

Annexe : • Tableau récapitulatif concernant la lecture (élaboré à partir de divers articles traitant des méthodes de lecture, des processus d'apprentissage dans ce domaine, etc.).

Annexe 1 :

- Notes récapitulatives concernant la lecture (élaborées à partir de divers articles traitant des méthodes de lecture, des processus et des conditions d'apprentissage dans ce domaine, etc.)



Si les théories linguistiques font prendre conscience de la distinction entre expression et communication, si elles sont à la base des oppositions de méthodes privilégiant soit l'oral, soit l'écrit, ou même recommandant les deux approches à la fois, on peut se demander, en ce qui concerne les processus et les conditions d'apprentissage de la lecture, c'est-à-dire d'un point de vue plus proche de la pratique des enseignants, si d'autres oppositions, qui furent nettement moins discutées et moins polémiques à l'époque, ne soulèvent pas des pistes de recherches et des problèmes plus pertinents.

Dans le tableau ci-après on peut résumer les oppositions qui, si elles ne sont pas encore entièrement levées aujourd'hui, révèlent au moins toute la complexité des processus et des conditions d'apprentissage de la lecture :

<u>AXE NON-SYMETRIQUE</u>			
	Encodage	Décodage	
De l'oral : le connu	parler	entendre/ écouter !?	<u>AXE DE DÉCALAGES</u>
Vers l'écrit : l'inconnu	écrire	lire	<u>DES ACQUISITIONS</u>

A ce tableau, on pourrait ajouter la distinction entre communication et expression qui soulève les nombreux problèmes de la signification (individuelle, puis sociale (convention reconnue)), de l'absence de signification (dans l'expression pure, par exemple) et du degré

d'abstraction dès que l'on utilise le signe plutôt que le symbole. Ces éléments, très proches de la linguistique, interrogent directement l'action pédagogique.

- Axe non-symétrique : on s'entend à dire que la lecture n'est pas l'opération inverse de l'écriture, plus généralement, l'encodage n'est pas l'opposé du codage et inversement.

- Par exemple on ne lit pas lettre par lettre, comme on peut le faire lorsqu'on ne fait que transcrire (exercice de copie) un message déjà codé. En bref, les pures correspondances entre sons et lettres, entre phonèmes et graphèmes, ou inversement, sont souvent non pertinentes.

Lors de cet exercice d'encodage, la construction peut se fonder sur l'intuition d'ensemble du contenu à coder (signification donnée) et la synthèse des parties (lettre par lettre par exemple) s'effectue élément par élément sur la trame fournie par cette intuition.

Si on décode par éléments successifs, chaque élément renvoie à un phonème, mais il y manquerait alors la signification.

- Ces remarques nous conduisent également aux oppositions entre les méthodes de type synthétique (allant du simple au complexe par combinaison d'éléments : lettres, mots, phrases) et les méthodes dites globales (partant de la saisie de structures graphiques complexes (avec l'énoncé global de leur signification) puis, par comparaison à d'autres structures analogues, allant à la découverte de structures graphiques partielles identiques : phrases, mots, lettres.).

- toutes deux cependant mettraient l'accent sur le décodage au détriment de l'encodage en présupposant que l'on sait d'autant mieux écrire que l'on sait lire.

- d'autres méthodes proposent de partir de l'encodage (avec comme point de départ une correspondance phonèmes/graphèmes !) pour aller vers le décodage.

On voit que les interrelations entre le global et le synthétique sont complexes en ce qui concerne les processus et les conditions d'apprentissage de la lecture et que l'on est loin de s'entendre à ce sujet.

- Axe de décalages des acquisitions : aux niveaux ontogénétique et phylogénétique, l'oral précède l'écrit.

- On propose alors de partir du connu, de l'oral, pour aller vers l'inconnu, l'écrit.

- D'un point de vue linguistique, certains auteurs insistent sur le fait que langue écrite et langue orale sont deux choses différentes. L'une repose davantage sur le symbole (le signifiant a un rapport direct avec le signifié), l'autre sur le signe (ici, le signifiant n'a plus de rapport direct avec le signifié; le rapport effectué résulte d'une convention arbitraire).

RECHERCHE-LECTURE : THÈMES DE DISCUSSION.

INTRODUCTION :

- **Présentation des thèmes discutés.**
- **Constestation de la méthode expérimentale.**

1) LA MÉTHODOLOGIE DE LA RECHERCHE :

LES PRINCIPALES INSUFFISANCES :

- **En ce qui concerne la comparaison des méthodes.**
- **En ce qui concerne la publication des résultats.**

- **Pertinence.**
- **Problèmes d'interprétation.**
- **En ce qui concerne l'utilisation ultérieure des résultats.**
- **Les bases théoriques de la "méthode de lecture" choisie ultérieurement par Maîtrise du Français.**

II) LE CONTEXTE HISTORIQUE ET SON INFLUENCE.

LE CONTEXTE POLITIQUE : LES ENJEUX, LES POLÉMIQUES.

- En pédagogie.
- En psychologie.
- En sociologie.

III) DE L'UTILITÉ EFFECTIVE DE LA RECHERCHE DE L'IRDP.

EN CONCLUSION :

- La méthode systémique comparative.
- Approche critique générale de cette recherche-évaluation.

**CH. 6D DISCUSSION DE LA RECHERCHE MATH
THÈMES DE LA DISCUSSION¹**

1) PREMIÈRE PARTIE .

Le nouveau curriculum de mathématiques et ses fondements psychopédagogiques

2) DEUXIÈME PARTIE :

Questions de méthodologie de l'évaluation

3) TROISIÈME PARTIE :

Questions d'épistémologie de l'évaluation (en référence à la typologie des modèles d'évaluation du cours)

4) QUATRIÈME PARTIE :

Questions de politique scolaire

5) CINQUIÈME PARTIE :

Le rôle de la recherche.

PREMIÈRE PARTIE

**LE NOUVEAU CURRICULUM DE MATHÉMATIQUES ET
SES FONDEMENTS PSYCHOPÉDAGOGIQUES**

RAPPELS :

- Le nouveau curriculum propose
de nouveaux objectifs
de nouveaux contenus,
de nouveaux matériels et de nouveaux moyens
d'enseignement.
- Il propose aussi, et de façon moins explicite ou indirectement, une
nouvelle pédagogie ou de nouvelles méthodes pédagogiques.

¹ En style normal, les éléments de réponse, les explications supplémentaires et le discours prévu. En italique, les questions posées.

- Au-delà des aspects mathématiques que nous n'aborderons pas et des aspects pédagogiques, le nouveau curriculum sert également la politique de coordination des systèmes cantonaux.
- En fait, le curriculum de mathématiques inaugure cette volonté.
- Son introduction veut s'effectuer dans une perspective de corrections et de rajustements successifs.

Par exemple, l'imprécision de l'énoncé des objectifs du nouveau curriculum de mathématiques devait permettre de ne pas fixer définitivement l'innovation et ceci pour affiner les nouveaux objectifs au fur et à mesure de l'expérience. Une des volontés des innovateurs était donc d'introduire le nouveau curriculum en le soumettant à l'épreuve de la pratique !

LES FONDEMENTS PSYCHOPÉDAGOGIQUES : QUELQUES OPTIONS

Parmi les données psychopédagogiques on peut citer les options suivantes :

- * une construction progressive et personnelle des connaissances;
 - * le primat de la compréhension sur le drill (conditionnement et mémoire);
 - * une conception cyclique du programme, conception qui s'oppose au principe d'une succession d'acquisitions additives propre à la pédagogie traditionnelle;
 - * "agir pour abstraire" : où l'action ici n'est pas à confondre avec la simple manipulation sensorielle et motrice;
 - * des activités de recherche (pédagogie de la découverte);
 - * travail par groupes (groupe ici se différencie du collectif ou de l'ensemble classe et de l'individuel).
- On remarque donc qu'il s'agit bien plus que d'un simple changement de contenu de programme et que d'un simple changement d'objectifs, mais d'une véritable révolution pédagogique. En effet, selon ces nouvelles options,
 - * il y a modification du rapport de l'élève au savoir;
 - * il y a modification du rapport de l'élève au maître;
 - * il y a même le risque de modifier les rapports internes et externes de l'institution-école elle-même.

- *Ce risque faisait-il partie des volontés des décideurs de l'époque ?*

Il n'y a jamais eu en Suisse de "mouvement révolutionnaire". Par contre, chez certaines des personnes qui ont préparé l'innovation, il est possible qu'il y ait eu une volonté de changement, d'améliorer les choses et de lutter contre le déterminisme social.

Cette forme d'idéalisme est présente depuis le 19^{ème} siècle. La fin des années 60 est marquée par une lutte idéologique intense entre la gauche et la droite. Les mouvements de rénovation sociale ont percé dans les nouveaux programmes. Pour les percevoir, il faut un oeil exercé. Le souffle révolutionnaire ne pouvait pas être exprimé trop clairement par crainte de troubles politiques. Il a fallu utiliser des termes adroits et appropriés afin de ménager les susceptibilités.

- *Ce nouveau curriculum n'était-il trop ambitieux en proposant autant de nouveautés à la fois ?*

On pensait que c'était une occasion unique de faire du neuf en mariant l'enthousiasme généreux de la gauche et la rationalité technocratique de la droite qui voulait mettre de l'ordre dans les programmes.

DEUXIÈME PARTIE

QUESTIONS DE MÉTHODOLOGIE DE L'ÉVALUATION

VERS UNE NOUVELLE CONCEPTION DE L'ÉVALUATION :

- On abandonne la comparaison de deux ou plusieurs méthodes car ce type d'approche :
 - * ne peut classer sur une même échelle des didactiques dotées d'objectifs différents;
 - * restreint le domaine d'études : toutes les variables ne peuvent être prises en compte (dans le réel, la variable indépendante n'existe pas);
 - * amène à des conclusions fragiles : on ne peut généraliser les résultats à d'autres situations.
- On va donc assister à un déplacement du regard dans l'évaluation des innovations pédagogiques :
 - si le regard portait avant tout d'abord sur la mesure des effets d'une innovation en termes de connaissances acquises chez les élèves;

- l'étude évaluative se centrera ensuite davantage sur la manière dont le projet innovateur est mis en oeuvre ainsi que sur ses modalités de réalisation dans son contexte institutionnel et social.
- Ce nouveau regard se rapproche de celui que l'on étudiera avec l'évaluation du nouveau curriculum de français (**recherche-action**).
- Dans l'évaluation du curriculum de français, il ne s'agit plus de se centrer sur les résultats des élèves, mais de mettre en place et de corriger le fonctionnement de l'innovation introduite. Dans l'évaluation du curriculum de mathématiques, on avait déjà réalisé partiellement cette même volonté. Ceci signifie que l'on abandonne la comparaison et que la méthode utilisée se veut de type non-comparatif.
- Pourtant, dans l'évaluation du nouveau curriculum de mathématiques, on a tout de même tenté de faire des comparaisons avec l'ancien curriculum.

Pourquoi ce retour vers l'ancien regard ?

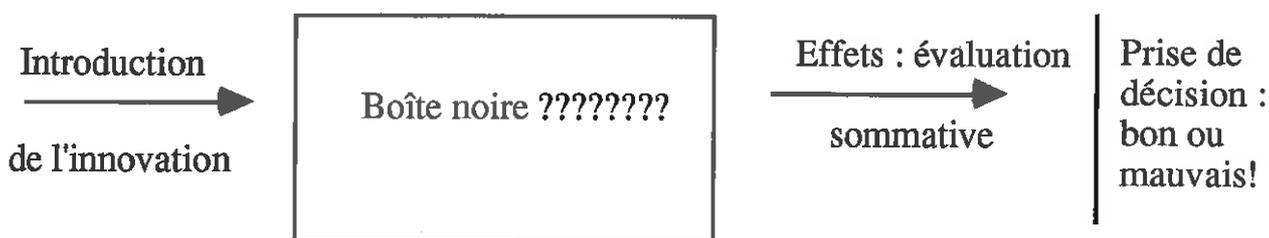
En partie, c'est la marque d'une inconséquence dans la méthode, mais c'est aussi un désir de combiner deux approches complémentaires, l'une qui mesure le résultat (quantitatif), l'autre qui l'explique (approche qualitative).

Quels ont été les principaux résultats de cette comparaison ?

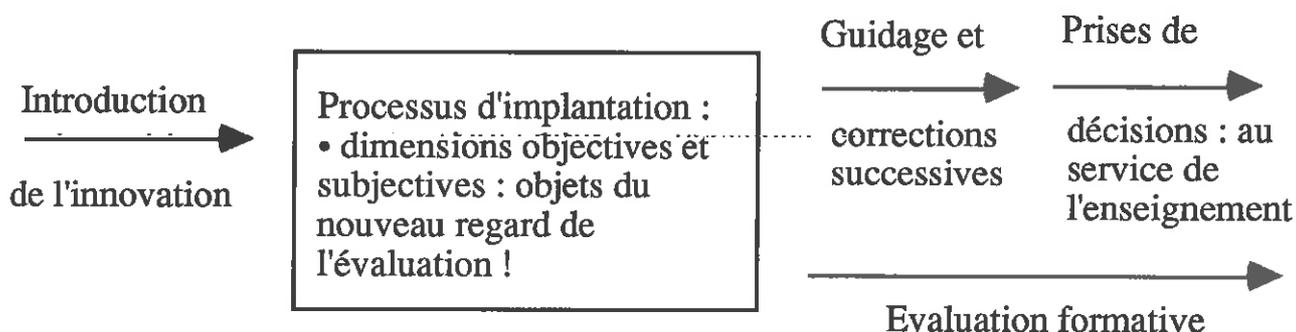
Le nouveau curriculum obtient les mêmes résultats que l'ancien en général (sur les objectifs communs), mais chacun l'emporte sur ses objectifs spécifiques. Mais l'éventail des questions abordées est plus large dans le nouveau programme.

POUR MIEUX SITUER CETTE NOUVELLE CONCEPTION DE L'ÉVALUATION, ON PEUT TENTER DE RÉSUMER CE QUI VIENT D'ÊTRE DIT DANS LES DEUX SCHÉMAS SUIVANTS :

A) ANCIEN REGARD



Exemple : la recherche-lecture.

B) NOUVEAU REGARD

Exemples : l'évaluation du nouveau curriculum de mathématiques (en partie du moins). L'évaluation du nouveau curriculum de français (que l'on étudiera plus tard).

COMMENTAIRES DES SCHÉMAS :

- On cherche à prendre en compte les multiples dimensions objectives et subjectives qui façonnent l'implantation d'une innovation et en influencent le cours.
- Ces dimensions devaient permettre de mieux comprendre les causes des effets obtenus.

En quelque sorte, on ouvre la boîte noire que représente le processus d'implantation;

et d'une évaluation sommative fondée sur un jugement du type bon ou mauvais, on passe à une prise de décision visant à guider l'innovation, la finalité pratique de la recherche est au service de l'enseignement.

QUELQUES CONSÉQUENCES MÉTHODOLOGIQUES DE CETTE NOUVELLE CONCEPTION DE L'ÉVALUATION

- Dans l'évaluation des mathématiques, on propose des recherches plus monographiques qu'expérimentales.

Comment peut-on définir une recherche monographique, quels en sont les principes généraux ?

Il s'agit d'analyser finement une situation particulière, en essayant d'en apercevoir la cohérence interne, d'en expliquer l'évolution par des mises en rapport aussi bien historiques que scientifiques.

Ceci est-il une conséquence directe de la nouvelle conception de l'évaluation ?

Oui, il faut prendre en compte tous les facteurs que les décideurs auront à considérer. Dans la recherche expérimentale, on recherche des lois générales. Dans l'approche monographique, on se situe dans le concret, dans une réalité particulière. Il s'agit de prélever suffisamment d'informations et de détails pour essayer de comprendre les relations entre les différents aspects. On s'intéresse à tout ce qui permet de saisir la logique interne des traits (comme un historien). Par exemple, dans l'évaluation du nouveau curriculum de mathématiques, on a introduit la récolte de données qualitatives, en interrogeant des philosophes et des logiciens.

Cette récolte est-elle nécessairement liée à la nouvelle conception de l'évaluation ?

L'évaluation sommative ne s'intéresse qu'aux résultats. L'évaluation formative s'intéresse aux processus, à tout ce qui influence les résultats pour agir sur les causes. Elle vise à comprendre les points de vue subjectifs et requiert donc une approche qualitative.

Comment résoudre le problème du caractère contradictoire des informations recueillies ?

C'était dans la stratégie de la recherche d'avoir l'occasion de trouver des résultats contradictoires (une des caractéristiques des recherches qualitatives). Plusieurs questions ont été posées plusieurs fois de façons différentes afin de faire des observations qui se recoupent. Si des contradictions apparaissent, on peut se demander s'il s'agit d'un effet de la formulation des questions. On peut aussi faire des hypothèses interprétatives (bon sens, connaissance du milieu, de l'histoire, de l'avis des enseignants) qui ne sont pas des réponses scientifiques, mais des hypothèses pour guider les corrections.

La CEM (commission de spécialistes, d'enseignants, ...) a été constituée afin d'assimiler et de tirer des conclusions de tous les éléments recueillis et de faire des propositions. En fait, la procédure s'est révélée assez floue (influence d'un leader, décision pas toujours la meilleure).

- On constate l'étroit parallélisme entre la nouvelle conception de l'évaluation elle-même et la nouvelle conception sur la manière d'introduire le nouveau curriculum.

Peut-on distinguer clairement ici entre innovateurs, décideurs et évaluateurs ?

Non, ils avaient ici partie liée du fait de la conduite de l'évaluation par l'IRD. Du point de vue de l'objectivité de l'évaluation, c'était une faiblesse. Du point de vue de la gestion de l'innovation, c'était au contraire un avantage.

Y a-t-il eu accord commun sur ces nouvelles conceptions ?

La nouvelle conception de l'introduction du curriculum et la nouvelle conception de l'évaluation de ce même curriculum sont-elles nécessairement liées ? Plus simplement, l'une ne va-t-elle pas sans l'autre ?

La prise en compte de tous les facteurs est une obligation pour les décideurs qui introduisent un nouveau curriculum, mais les évaluateurs ont longtemps voulu porter un jugement sur l'essence de l'innovation en la distinguant de tous les accidents qui pouvaient perturber plus ou moins son fonctionnement.

• *Cette nouvelle approche ne demande-t-elle pas une recherche sur la légitimité des nouveaux objectifs ?*

Effectivement, il devient nécessaire alors de distinguer la valeur des objectifs poursuivis, de la valeur des résultats obtenus (contrairement à l'approche de l'évaluation par objectifs de Tyler)

LES QUATRE VOLETS DE L'ÉVALUATION DU NOUVEAU CURRICULUM DE MATHÉMATIQUES

- 1) Enquêtes par questionnaire auprès des enseignants
- 2) Enquêtes par groupes sur les moyens d'enseignement (groupes composés d'enseignants et de formateurs)
- 3) Les recherches psychopédagogiques
- 4) Les tests de connaissances (surveys)

1) ENQUÊTES PAR QUESTIONNAIRE AUPRÈS DES ENSEIGNANTS

a) But :

- * sonder l'accueil fait au nouveau curriculum
- * connaître le point de vue des enseignants sur les différentes composantes de l'innovation (les nouveaux objectifs, les méthodes, le travail par groupe, les moyens d'enseignement, etc.)

Question : on a bien les opinions des enseignants, mais pas les observations réelles faites au sein des classes. Comment combler le décalage entre intentions et pratiques réelles ?

Si la recherche avait été menée dans l'optique objectiviste, il aurait fallu faire une observation dans les classes. Cela n'a pas été fait ici, un peu par crainte de la réaction négative des enseignants, mais aussi parce que le problème n'était pas là. Il ne s'agissait pas de rechercher les causes ultimes, mais de savoir comment le nouveau curriculum était accueilli par les enseignants, les parents. Le souci des décideurs était d'éviter la dérive de l'innovation, ou des conflits majeurs. L'avis des enseignants suffisait donc aux autorités, mais on peut se demander alors si la recherche effectuée est une véritable recherche ou une assurance tout risque.

b) Les multiples aspects de la participation dans l'évaluation :

- *L'entrée d'une partie des enseignants dans l'évaluation a-t-elle servi indirectement à la formation de ces mêmes enseignants, par exemple :*
 - *rappeler quels étaient les objectifs du curriculum (en faire prendre conscience aux enseignants, lorsqu'ils devaient répondre au questionnaire);
 - * éclairer et sensibiliser les enseignants sur la démarche des enfants lorsqu'ils devaient protocoler les réponses de ceux-ci (concerne les surveys)
- On voit ainsi que cette participation des enseignants à l'évaluation intervenait directement sur la réalité.

Quelles sont les limites de cette participation, ou de cette façon de procéder ? (Manipulation)

Cet aspect a été développé par les physiciens en 1920 : on ne peut pas observer une particule sans l'influencer. Le regard modifie la réalité. Pour cette recherche, on a demandé l'avis des enseignants sur chacun des objectifs du nouveau curriculum. Interroger tous les enseignants contribuait à leur formation continue (en leur faisant "réviser leurs leçons"). Il y avait donc bien un aspect manipulateur, puisqu'on voulait intervenir sur le comportement des enseignants.

Maintenant, est-on suffisamment clair lorsque l'on parle de participation ?

Il faut en effet distinguer entre une participation implicite, sans que les intéressés soient forcément conscients de leur intégration à la recherche, et une participation explicite, dans laquelle les intéressés sont pleinement informés des attentes formulées à leur égard.

• Mais est-ce là la signification principale du caractère participatif dans l'évaluation ?

Non, il ne faudrait parler de participation que lorsque les intéressés prennent directement part à l'évaluation : par exemple, dans les cas où leurs avis sont pris en compte et font partie intégrante de l'évaluation ?

• Finalement, dans quelle mesure et comment réaliser la participation des partenaires autres que les décideurs ?

C'est un gros problème, car la participation de représentants des enseignants ou des parents ne touche que très partiellement les milieux correspondants. Il faut trouver des opérations où ils soient personnellement impliqués.

En ce qui concerne le nouveau curriculum de mathématiques, quel a été le degré de participation des enseignants ?

Dans la gestion des entreprises, la participation est la possibilité offerte aux collaborateurs d'agir et d'intervenir sur les décisions. Dans la recherche math, la participation des enseignants n'a pas été de cette nature. L'avis des enseignants a été considéré de manière subjective.

d) Apport des enquêtes par questionnaires :

Le caractère officiel et non anonyme des questionnaires a-t-il eu pour effet d'augmenter les réponses favorables à l'innovation ?

* On relève l'attitude positive des enseignants (environ les 2/3), mais "dans toutes ces déclarations, il est bien difficile de faire la part entre ce qui relève d'une réelle conviction éprouvée ou d'un acquiescement conformiste aux nouvelles perspectives pédagogiques officielles" !

Beaucoup d'indices font en effet penser que les résultats recherchés n'étaient pas atteints. On retrouve avec le nouveau curriculum le même formalisme qu'auparavant; il suffit que l'école touche quelque chose pour que l'activité se sclérose !

On relève malgré l'enthousiasme général, la persistance de l'application des normes traditionnelles, notamment en l'absence de normes nouvelles plus explicites et claires (cf. les opérations).

On souligne également des difficultés comme le décalage qui subsiste entre les intentions initiales fondées sur les dernières théories de l'apprentissage et la généralisation du projet dans la réalité quotidienne de la classe.

Le Service de Recherche souhaitait préserver l'anonymat. Ce sont les cantons qui ont demandé que les questionnaires ne soient pas anonymes. Deux tiers des chefs de service ne supportaient pas l'anonymat car ils l'assimilaient à des lettres anonymes. Cet argument cachait peut-être la volonté de faire pression sur les enseignants pour obtenir des résultats positifs et faire taire les opposants. C'est un problème déontologique. Pour la liberté d'expression et d'opinion, il est indispensable de respecter l'anonymat.

Ces observations impliquent-elles la non-pertinence des réponses recueillies dans les questionnaires adressés aux enseignants ?

Non, dans ce cas particulier, le non-respect de l'anonymat ne remet pas en cause les conclusions d'ensemble, mais seulement les pourcentages de réponses positives. Par ailleurs, un certain nombre de non-réponses aurait pu être interprété comme des réponses négatives. Toutefois, plusieurs éléments ont montré que les résultats obtenus étaient assez proches de la réalité et on ne pouvait pas ne pas publier les résultats.

2) ENQUÊTES PAR GROUPES SUR LES MOYENS D'ENSEIGNEMENT (GROUPES COMPOSÉS D'ENSEIGNANTS ET DE FORMATEURS) :

a) Perspective :

- * d'adaptation et d'assimilation des méthodologies nouvelles, notamment en faisant participer les enseignants.

b) But :

- * confronter la pertinence du nouveau programme au terrain pédagogique.

c) Apport des enquêtes par groupe .

- * L'un des principaux apports de ces enquêtes est de montrer que, même s'il y a consensus sur les diagnostics (aussi bien sur les insuffisances que sur les traits positifs), il y a plusieurs manières de modifier les moyens d'enseignement : en clair les choix et les décisions pédagogiques à prendre par la suite sont multiples !

Quelles conclusions peut-on tirer de cet apport ?

De ces enquêtes sont ressortis des faits sur lesquels il était possible de s'entendre. Tout le monde voyait les mêmes problèmes, mais pas les mêmes solutions (par exemple pour les bases : abandonner ou continuer avec une méthodologie plus efficace).

Une commission (méthodologues, mathématiciens, ...) a été créée pour élaborer des solutions et trouver un consensus hors de l'officialité d'un milieu politique.

Ce type d'enquêtes n'était-il pas alors spécialement adapté à la nouvelle conception de l'évaluation ?

En effet, c'est par une discussion des différentes interprétations possibles des résultats observés que l'on peut préparer des décisions consensuelles.

3) LES RECHERCHES PSYCHOPÉDAGOGIQUES :

* Suite aux problèmes pédagogiques dégagés dans les enquêtes précédentes, on a mené des études centrées sur des points très précis : par exemple l'étude sur l'introduction de la soustraction en 1P; l'étude sur la compréhension des énoncés de problèmes et la distinction de trois étapes dans la résolution de problème : prise de connaissance du problème, compétences à mettre en oeuvre, et procédures de résolution du problème, cette dernière étape étant assez bien étudiée en psychologie cognitive, contrairement aux deux autres; la construction du nombre chez l'enfant; etc.

Ces recherches ont-elles été intégrées dans les modifications apportées au curriculum par la suite ? (ex. de la construction du nombre)

Les recherches psychopédagogiques ont aidé à la compréhension des problèmes, mais n'ont pas été utilisées directement pour modifier le curriculum. Elles ont apporté des documents de base qui ont permis de faire des propositions aux décideurs. Elles ont aussi permis d'accumuler des informations de portée plus générale. Elles étaient donc importantes pour le développement de la didactique. On retrouve là les deux pôles (appliqué et fondamental) que devrait contenir toute recherche. Se situer à un extrême ou à l'autre n'est pas une bonne solution.

4) LES TESTS DE CONNAISSANCES (SURVEYS) :

* On a souligné l'insuffisance des opinions recueillies par les questionnaires et l'on s'est posé la question de savoir si les nouveaux objectifs annoncés avaient été atteints ou non.

Pourquoi se poser une telle question dans les perspectives d'évaluation et d'introduction du nouveau curriculum qui avaient été adoptées ?

Deux positions méthodologiques étaient présentes : une étude objective (Jean Cardinet) et une évaluation plus formative (Jacques Weiss). Ces deux positions ne s'opposent pas. Les tests de connaissance (objectifs) et les questionnaires (qualitatifs) ont permis d'aborder les mêmes problèmes sous des angles différents (triangulation). Il faut confronter les avis des enseignants entre eux et avec les résultats individuels.

D'un point de vue politique, les DIP avaient intérêt à publier les résultats obtenus : le nouveau curriculum marche, en tout cas aussi bien qu'avant.

- Les surveys rassemblent trois types de tests :

Rappels :

a) les tests collectifs

Les élèves d'une même classe répondent à des questions différentes et ces questions varient également d'une classe à l'autre.

Le nombre élevé d'élèves permet d'élargir le champ des questions en demandant un effort minimum à chaque enfant.

L'échantillon (1200 classes, pour 20000 élèves environ (1P)) est suffisamment grand pour permettre d'estimer précisément le taux de réussite à chaque objectif choisi et de dégager en même temps l'effet d'autres sources de difficultés dans la résolution de problèmes.

b) les tests individuels du genre clinique

Les épreuves individuelles du type clinique passées par les enseignants leur permettent d'apprendre à se mettre à l'écoute de leurs élèves.

c) et les tests comparatifs (entre l'ancien et le nouveau curriculum).

Effectués en fin de 4P pour les premiers.

Questions :

- Dans les tests de connaissances appliqués aux élèves de chaque année, on n'observe pas de test de fidélité. Les chercheurs ont même fait varier intentionnellement les questions sur un même sujet.

Pourquoi ?

La psychologie différentielle a pour objet de différencier les élèves. Pour évaluer la fidélité des tests, on utilise un critère de référence (note du maître, note globale au test). On garde les questions qui ont une bonne corrélation avec ce critère. Cette manière de procéder n'est valable que si on veut classer les élèves de bons à mauvais. Dans la recherche math, ce n'était pas le but poursuivi. On voulait contrôler la valeur des objectifs introduits dans le curriculum et savoir s'ils étaient atteints ou non. Il fallait donc différencier les objectifs et non pas les élèves. Les questions proposées devaient très différentes et graduées pour savoir jusqu'où le curriculum avait atteint ses buts. La méthode des psychologues procède de manière opposée : rejet des questions réussies ou ratées par tout le monde. L'évaluation du curriculum nécessitait une méthode appropriée, différente de celle des psychologues.

- Contrairement aux enquêtes faites dans les autres pays, on a introduit des interviews cliniques dans les tests et dans les questionnaires.

Pourquoi ?

Le nouveau curriculum se devait de tenir compte de l'approche piagétienne. Il fallait pouvoir analyser le raisonnement des élèves et les sources de difficulté à écarter. Un simple constat d'échec ne permettait pas d'orienter les corrections.

TROISIÈME PARTIEQUESTIONS D'ÉPISTÉMOLOGIE DE L'ÉVALUATIONCOMMENT SITUER L'ÉVALUATION DU NOUVEAU CURRICULUM DE MATHÉMATIQUES DANS LA TYPOLOGIE DES MODÈLES D'ÉVALUATION DU COURS ?ÉPISTÉMOLOGIE :

- (Utiliser transparent du cours !)
- La nouvelle conception de l'évaluation ne modifie-t-elle pas les finalités mêmes de l'évaluation ?

En d'autres termes, au fur et à mesure de l'évolution des méthodes d'évaluation, ne serait-on pas contraint de remettre en cause les buts mêmes de l'évaluation ?

Dans notre cas, un des buts de l'évaluation du nouveau curriculum était d'assurer la réussite de son introduction, car la nouvelle conception de l'évaluation ne contenait pas la possibilité de supprimer le nouveau curriculum; cela aurait été tout à fait opposé à la demande des décideurs.

[Exemple de l'individu qui, lors de sa formation, doit reconsidérer ses premiers buts. Ceux-ci deviennent même parfois caducs !]

- Une autre conséquence de la nouvelle conception de l'évaluation concerne l'objet de l'évaluation : on se centre plus sur les processus en cours que sur les effets.

Dans le cas des mathématiques, *une telle conception de l'évaluation ne conduisait-elle pas nécessairement à ne modifier le nouveau curriculum que sur des points mineurs ? (Cette prudente réserve semblait d'ailleurs correspondre à la demande des décideurs).*

Les premiers travaux d'évaluation de curriculum (au milieu de l'école) étaient essentiellement sommatifs. Si les résultats n'étaient pas satisfaisants, il était logique d'abandonner les nouveaux programmes (par exemple aux Etats Unis, les programmes commerciaux offerts "en paquet" aux écoles).

Les travaux d'évaluation plus récents cherchent à comprendre le pourquoi de ce qu'on observe et à saisir la continuité par une démarche plus qualitative. Dans ce cas, il est plus difficile d'accepter ou de rejeter globalement les innovations. Il s'agit plutôt de recueillir des informations pour pouvoir modifier des points particuliers, plus ou moins importants, du nouveau curriculum.

Dans cette recherche, étant donnés les investissements financiers et personnels, il était impensable de tout rejeter en bloc. D'un autre côté, comme rien n'est parfait, on ne pouvait pas dire simplement "tout va bien". Les chercheurs étaient effectivement condamnés à apporter des modifications plus ou moins mineures au curriculum.

L'évolution des objectifs était pourtant présente dans l'esprit des chercheurs. Si les analyses réalisées n'ont pas été directement appliquées, elles ont néanmoins conduit à repenser les objectifs.

En faisant une recherche sur "comment l'enfant apprend", on en est venu à définir autrement le sens de la mathématique. La conception actuelle met l'accent sur le caractère constructif de la mathématique et sur l'activité des élèves. L'enfant doit apprendre à mathématiser, à construire lui-même, à voir de l'abstrait dans du concret.

La conception de l'évaluation formative a-t-elle changé de façon importante depuis l'époque où fut introduit le nouveau curriculum de mathématiques jusqu'à nos jours ?

La conception de l'évaluation formative a effectivement évolué. Au départ, l'évaluation était perçue comme un jugement. La plupart des acteurs pensaient que les chercheurs allaient conclure que le curriculum était bon ou mauvais. Le formatif venait comme une conséquence ultérieure du sommatif (boucle de réparation et de remédiation). C'est pourquoi les autorités n'étaient pas d'accord de faire des modifications avant que la totalité des degrés aient été évalués (évaluation formative rétroactive). Par la suite, on a pris conscience qu'il ne suffisait pas d'observer les résultats, mais qu'il fallait comprendre qualitativement ce qui se passe dans la tête des élèves. L'évaluation s'est alors centrée sur les processus.

Actuellement, l'évaluation formative consiste à voir comment les enseignants enseignent, comment les enfants apprennent. A ce niveau, on peut observer déjà certains phénomènes sans recourir à du quantitatif. On peut intervenir à ce moment déjà (évaluation formative interactive)

En résumé, le nouveau regard jeté sur l'évaluation considère l'introduction du projet comme un processus de changement (changement des représentations notamment) et d'une appropriation du projet lui-même.

UTILITÉ :

Quelle échelle d'utilité a servi de référentiel ? Les instances de décision n'ont-elles pas gardé le contrôle des opérations d'évaluation, les chercheurs tentent vainement de s'intéresser à des utilités particulières, mais sans grand succès ?

- Il est vrai qu'ici l'utilité supposée commune a été l'utilité estimée par les instances de décision, c'est-à-dire les autorités scolaires.

En cours d'évaluation, les chercheurs ont impliqué d'autres instances (parents, enseignants, opinion générale publique,...).

L'utilité n'était pas la même pour ces différents acteurs. Poser des questions aux enseignants témoignait de l'intérêt qu'on portait à leur avis. Mais leur point de vue n'était pas le même que celui des méthodologues, des parents, des autorités, ... Dans l'esprit des chercheurs, la multiplicité des échelles de point de vue était importante.

A priori, il y avait donc des utilités particulières. Cependant, en dernier ressort, ce sont les autorités qui ont pris les décisions. Il s'agit alors d'une échelle d'utilité unique. Les utilités particulières permettent seulement aux autorités de tenir compte des différents avis dans leur stratégie politique (rester au pouvoir).

ORIENTATION :

- *L'orientation de cette évaluation va, il est certain, vers la décision, mais de quel type de décision est-il question ? S'agit-il de faire mieux respecter la discipline des programmes imposés, ou de modifier les programmes initiaux pour satisfaire les critiques ?*

La recherche peut effectivement parfois servir à la justification de décisions déjà prises. Pour arriver à faire fonctionner le système scolaire, les autorités utilisent les résultats pour éviter les vagues (minimiser les problèmes, calmer les esprits, redonner confiance aux enseignants). Corriger uniquement les problèmes mineurs peut relever d'une telle pratique.

MÉTHODE :

On a déjà parlé de quelques aspects méthodologiques, notamment de l'approche comparative.

- On rappelle que Tyler défendait l'idée "qu'il suffisait, pour valider un curriculum, de prouver qu'il permettait d'atteindre les objectifs éducatifs visés." (Cf. Évaluation opérationnelle non comparative, page 5, chapitre 3 du cours.)
- Cette idée, on le voit, c'est-à-dire regarder les résultats des élèves en regardant si les objectifs sont atteints, subsiste encore dans l'évaluation du nouveau curriculum romand des mathématiques.

Ce point de vue est de type objectiviste et va à l'encontre des points de vue exprimés précédemment : où peut-on alors situer l'évaluation du curriculum de mathématiques dans cette typologie ?

On ne peut situer précisément la recherche math dans la typologie proposée lors du cours. Méthodologiquement, elle comporte des aspects expérimentaux (tests comparatifs) et d'autres aspects plus proches de la recherche action (former les enseignants). A une démarche méthodologique pure, on a préféré l'utilisation d'un spectre de méthodes à l'intérieur de la recherche.

Les sciences sociales n'ont pas pour objet de mettre en évidence des lois généralisables à tous les lieux et tous les temps, mais à faire évoluer toute la société en amenant les gens à se poser des questions, à tâtonner. On avance tous en même temps grâce à la multiplicité des recherches et des réflexions qui modifient nos représentations.

QUATRIÈME PARTIE

QUESTIONS DE POLITIQUE SCOLAIRE

• SELON JAQUET (1978) :

“Toute remise en cause trop fondamentale de l’innovation est hors de question pour l’instant, ne serait-ce qu’à cause des répercussions qu’une telle critique aurait dans l’esprit du public et du corps enseignant. L’évaluation doit travailler dans le cadre précis d’un ajustement limité du curriculum et de sa gestion.”

“Les évaluateurs doivent donc apprendre à se situer par rapport à l’attente des autorités scolaires. Leur tâche ne consiste pas seulement à évaluer, mais aussi et surtout à s’intégrer dans la gestion de l’innovation et à faire réussir l’entreprise...”

“Mais dans le système adopté en Suisse romande, il convient aussi que les responsables de l’évaluation restent conscients de l’ensemble du système où elle s’insère et prennent garde à ce que ne soient pas suscitées des perturbations plus graves encore que celles qu’elle aperçoit et qu’elle doit aider à corriger.”

Quelles conclusions peut-on tirer de ces remarques ? Est-il encore possible de parler de “science” dans le cas des évaluations ?

- F. Jaquet fait également remarquer que le premier rapport sur la première année primaire est resté largement ignoré des décideurs !

La lenteur du processus d’évaluation et le manque de vue d’ensemble n’ont-ils pas contribué à perdre en cours de route les corrections et les adaptations régulatrices nécessaires ?

- Les chercheurs soulignent bien l’impossibilité de comparer l’ancien et le nouveau curriculum, notamment à cause d’objectifs fondamentalement différents, mais ils déclarent céder à la demande sociale. Ils décideront alors de faire porter la comparaison sur chaque domaine particulier et ils éviteront de procéder à une comparaison globale.

Dans le contexte politique de l’époque, fallait-il à tout prix un bilan positif pour justifier l’emploi du nouveau curriculum ?

Il n’y a pas de réponse évidente et indiscutable à ces questions. Ces critiques comportent une part de vérité, mais n’expriment pas toute la vérité non plus. Elles font réfléchir à des dangers réels qui menacent toujours les recherches d’évaluation.

CH. 7C. DISCUSSION DE L'ÉVALUATION DU CURRICULUM DE FRANÇAIS

SOMMAIRE

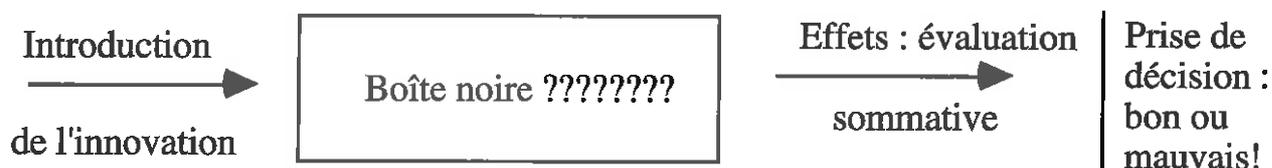
DÉPLACEMENT DU REGARD.....	1
a) Ancien regard.....	1
b) Nouveau regard.....	1
LES RECHERCHES FONDÉES SUR LA PARTICIPATION.....	2
1) L'observation participante.....	2
2) L'animation non-directive.....	2
3) La recherche-action.....	3
4) L'observation interactive.....	3
FONDEMENTS ÉPISTÉMOLOGIQUES.....	3
Caractéristiques.....	4
Quelles sont les principales conséquences de ces caractéristiques ?.....	4
En résumé.....	5
L'OBSERVATION INTERACTIVE.....	5
1) La démarche proposée.....	5
2) Les justifications d'une telle approche.....	7
LE PROBLÈME DE L'OBJECTIVITÉ.....	8
Distorsion de l'information : par quels mécanismes cela arrive-t-il ?.....	8
EN CONCLUSION.....	10
BIBLIOGRAPHIE.....	12

CH. 7C. DÉPLACEMENT DU REGARD

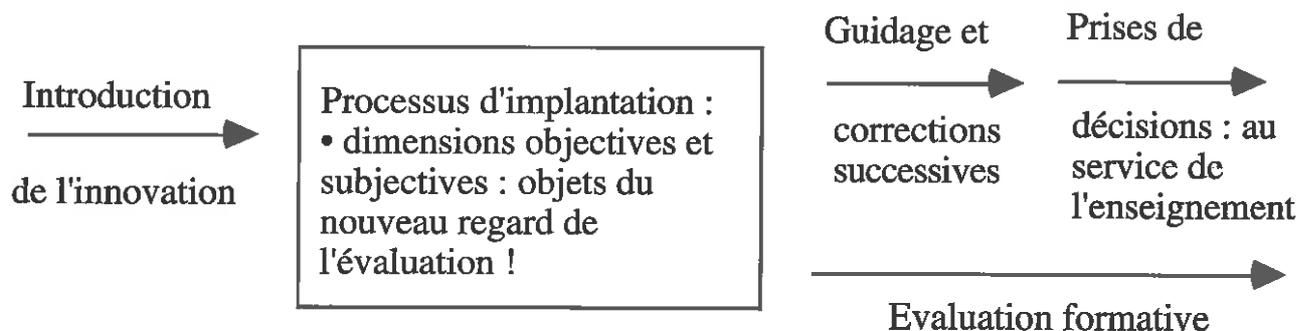
- Dans l'évaluation du nouveau curriculum de mathématiques, nous avons assisté à une première tentative, à une intention de déplacement du regard.

- Pour mieux situer cette nouvelle démarche, on avait examiné les deux schémas suivants :

A) ANCIEN REGARD



B) NOUVEAU REGARD



- Le second schéma montre que l'on cherche désormais à prendre en compte les multiples dimensions objectives et subjectives qui façonnent l'implantation d'une innovation et en influencent le cours.

- Il ne s'agit donc plus tellement de voir les connaissances acquises chez les élèves (évaluation sommative), mais de mettre en place et de corriger le fonctionnement de l'innovation introduite (évaluation formative).

- Ce nouveau regard ou cette nouvelle démarche évaluative caractérise assez bien ce qui sera théoriquement posé dans l'évaluation du nouveau curriculum de français.

- Aujourd'hui, et contrairement aux exposés précédents, mon objectif premier n'est pas de montrer comment et dans quelle mesure cette nouvelle démarche a été, ou n'a pas été, effectivement réalisée, mais de présenter les principes qui sont à la base de ce modèle d'évaluation.

- A cet effet, je parlerai successivement :

- 1) des recherches fondées sur la participation;
- 2) des fondements épistémologiques de ces recherches;
- 3) de l'observation interactive en particulier;
- 4) de l'analyse critique de Michael Huberman;

LES RECHERCHES FONDÉES SUR LA PARTICIPATION

- Nous avons déjà discuté de quelques aspects de la participation dans l'évaluation.

Par exemple, nous avons vu dans quelle mesure les chercheurs avaient impliqué les enseignants dans l'évaluation du nouveau curriculum de mathématiques. Dans ce cas, la participation des enseignants avait été indirecte et nous avons souligné le danger de manipulation.

- Avec les recherches dites "fondées sur la participation", c'est l'implication même des chercheurs qui s'accroît et non plus seulement celle des participants à l'action éducative étudiée.

- On peut donner à titre d'exemple quelques unes de ces démarches fondées sur la participation :

1) L'observation participante :

A l'origine, elle est décrite par Malinowski (anthropologue). Elle invite le chercheur à vivre la vie des groupes qu'il étudie et à partager ainsi leurs activités. L'idée est de permettre au chercheur de mieux comprendre la vision du monde des groupes qu'il observe et de pouvoir situer leurs discours dans un contexte socioculturel particulier. Le chercheur éviterait ainsi le culturo-centrisme et se distancierait de ses propres valeurs.

2) L'animation non-directive :

Ici, le chercheur joue un rôle dans le fonctionnement du groupe. Sans entrer en discussion sur le concept de "non-directivité", le chercheur va plus loin que le simple partage, il agit. Et puisqu'il agit, il porte une part de responsabilité dans les décisions qui sont prises et les conséquences qui

en découlent.

3) La recherche-action :

Théoriquement, dans la recherche-action le chercheur partage, agit et s'engage en faveur de valeurs et d'objectifs définis. Cet engagement aura des implications pédagogiques, psychologiques, sociales et idéologiques plus ou moins marquées : ici, en interagissant avec l'objet qu'il étudie, le chercheur participe directement à la transformation du réel.

En tant que scientifique, cependant, il doit rester le garant d'une certaine objectivité. Son engagement subjectif ne le dispense pas pour autant d'une réflexion : il lui faut faire la distinction entre son influence propre et celle de la réforme scolaire ou de tout autre objet qu'il évalue.

Ce modèle implique en général une distribution des rôles entre cadres administratifs et enseignants. Pensée et action, décision et exécution, ne sont plus attribuées exclusivement à certains groupes. Dans cet esprit de partage, il y a aussi l'idée d'un projet commun.

En résumé, la recherche-action ouvre à une compréhension du monde en termes de changements, de conflits et d'innovations. Elle se situe à la limite de la connaissance et de l'action.

4) L'observation interactive :

Il s'agit du modèle utilisé pour l'évaluation du nouveau curriculum de français. J'en donnerai les motivations par la suite.

Pour le moment, j'aimerais montrer que tous ces modèles se fondent sur des considérations d'ordre épistémologique.

FONDEMENTS ÉPISTÉMOLOGIQUES

Voici quelques caractéristiques fondamentales dont tous ces modèles essaient de tenir compte :

- Ils insistent d'abord sur le rôle de la prise de conscience chez les participants à une recherche.

Qu'il s'agisse d'enfants, d'enseignants, d'un établissement scolaire tout entier ou même de toute une institution, le chercheur ne se trouve pas en face d'objets inanimés ou dénués de conscience.

Les participants à l'expérience, à l'évaluation, ne sont pas passifs: ils réagissent à la manière dont on les traite, d'une façon qui fait immédiatement éclater le cadre limité dans lequel le chercheur ou l'évaluateur voudrait les maintenir.

Dans cette perspective, les éléments subjectifs peuvent devenir des déterminants essentiels ! La situation est dite réactive.

- La dissociation entre l'observateur et l'observé s'avère aussi difficile. En effet, à moins de renoncer à toute utilité pratique, ou au risque de laisser survenir une catastrophe, l'observateur qui s'aperçoit d'un problème est amené à intervenir pour le résoudre. Mais en réagissant, il s'implique lui-même dans la recherche !

- Enfin, selon que le chercheur se situe entre la connaissance et l'action, les finalités de la recherche sont différentes :

		<u>Versant choisi</u>	
		Régularités	Significations
But recherché	Connaître	R. modélisante	R. interprétative
	Agir	R. technologique	R. décisionnelle

Schéma tiré de J. Cardinet, L'évaluation du nouvel enseignement de la mathématique en première année primaire : présentation de l'expérience romande, Neuchâtel : IRDP/R 77.17, Août 1977.

Commentaires du schéma :

Les recherches de la colonne de gauche tentent de répondre à la question : quelles sont les causes ?

Les recherches de la colonne de droite tentent de répondre à la question : quelles sont les raisons ? Les raisons évoquent les intentions sous-jacentes aux faits qui se sont produits ou qui sont en train de se produire.

Généralement, ces dernières recherches tiennent compte du temps. Si le chercheur n'a aucune prise possible sur les faits du passé (recherche historique), l'évaluateur d'un contexte actuel peut au contraire contribuer à la modification des faits en cours de réalisation.

Quelles sont les principales conséquences de ces caractéristiques ?

- Tout d'abord, la situation réactive amène à considérer des éléments non objectifs comme : les significations, les valeurs, les projets et les intérêts des divers partenaires. La situation est comprise de l'intérieur par le chercheur.

- Du coup, la notion même de "réel" prend une acception différente : l'entrée d'éléments subjectifs ne permet plus d'aboutir dans la recherche à un réel unique, à une vérité unique.

- Les résultats ne sont pas indépendants des conditions d'observation

dans lesquelles ils ont été produits : le temps et le lieu de leur production débouchent sur des connaissances particulières qui tendent vers le concret et le local.

- Si l'on se centre sur les conduites concrètes et dans les conditions et dans les lieux dans lesquels elles émergent, on est exposé à leur complexité et à leur imprévisibilité, d'où l'obligation de se résigner à ce que leur sens ne soit jamais épuisé.

On n'épuise pas la réalité. L'exhaustivité n'est jamais atteinte, car le sens se recrée sans cesse.

En résumé et d'une manière générale, les sciences humaines doivent tenir compte de traits spécifiques propres à leurs champs d'études :

- "La prise de conscience d'un aspect de la réalité sociale modifie du même coup cette réalité elle-même. Toute intervention dans un système social, ne serait-ce que la présence d'un observateur, modifie la situation à l'intérieur de ce système."

- "Le monde social (celui des intentions et des significations) est perçu de l'intérieur, par une sorte d'identification à l'autre. A côté d'une causalité objective reliant des faits, on est en présence aussi d'une causalité subjective reliant des significations. On ne demande plus seulement "comment cela s'est-il produit ?" mais aussi "pourquoi est-ce arrivé ?"

Ceci amène naturellement les chercheurs à une prise en compte de la subjectivité :

- "Suite au fait que l'intervention modifie l'objet d'étude, que nous nous plaçons dans une situation réactive, une démarche réflexive est demandée au chercheur, qui doit intégrer dans son plan de recherche les effets de ses modes d'intervention."

- Toute intervention qui modifie l'état de choses existant paraît ainsi menaçante. C'est pourquoi, suite à ces constatations, on propose de constituer un projet commun et de gérer cette intervention de façon ouverte et participative, afin de faciliter l'acceptation et l'intégration du changement.

L'OBSERVATION INTERACTIVE

La démarche évaluative suivie, **une observation interactive**, est la démarche adoptée dans le cas de l'évaluation du nouveau curriculum de français.

1) LA DÉMARCHE PROPOSÉE :

- Pour éviter les désavantages d'une évaluation sommative, comme dans une recherche-expertise :

- l'évaluation vient trop tard
- elle est centrée sur ce que les experts considèrent comme important
- elle risque d'ignorer ce qui fait vraiment problème pour d'autres;

- les chercheurs proposent une approche qui tente :

- de coordonner les interventions des spécialistes, du début à la fin de l'évaluation;
- de tenir compte et de répondre efficacement aux remarques effectuées par les acteurs non-spécialistes (enseignants et parents);
- d'organiser la coopération des enseignants entre eux et avec les parents;
- d'assurer la circulation des informations à tous les niveaux du système scolaire.

- Si les chercheurs soulignent les désavantages d'une telle démarche :

- le flou du partage des responsabilités;
- la lourdeur des prises de décision;
- l'incertitude des plans à long terme;
- le coût (en temps et en argent) de ces multiples interactions.

- Ils en montrent également les avantages :

- il y a mise en évidence plus rapide des difficultés et possibilité de réagir à temps;
- l'évaluation est de type formatif: elle permet de corriger les points de détail, sans remettre en cause l'ensemble;
- l'observation reste ouverte, rien n'est exclu à l'avance;
- il y a exploitation des compétences respectives de chacun;
- il y aurait acceptation plus facile du changement.

- Pour les chercheurs, ce type d'approche :

- permet d'un côté la reconnaissance et la valorisation de l'expérience des enseignants;

- mais d'un autre côté, elle limite la participation des enseignants. En effet, elle n'est pas une recherche-action au sens décrit précédemment, car les principes et les objectifs de la recherche ne sont pas conçus de manière paritaire.
- Enfin, elle diffère de la recherche-action dans le sens où l'organe de gestion de l'évaluation n'a pas non plus la possibilité de remettre en cause ses statuts pour effectuer des adaptations au niveau même de ses principes de fonctionnement et de ses objectifs.

2) LES JUSTIFICATIONS D'UNE TELLE APPROCHE

Ce type d'approche est-il justifié ? En ce qui concerne l'introduction et l'évaluation de ce nouveau curriculum :

- les personnes impliquées sont bien confrontées à la nécessité d'un changement de leurs représentations, de leurs attitudes et de leurs comportements !

- Changer les conduites peut impliquer diverses réactions :

- réactions de défense (actives ou passives)
- réactions qui peuvent être relativement indépendantes de la qualité intrinsèque du curriculum

Nous sommes bien dans une situation dite réactive.

- Comme toute intervention dans une telle situation, on retrouve en ce qui concerne l'introduction et l'évaluation du nouveau curriculum de français quelques unes des caractéristiques citées auparavant :

- nous nous situons bien sur le plan du vécu, du subjectif, de l'interprétation des significations et de l'étude des processus;
- les chercheurs sont également amenés à prendre position par rapport à une échelle de valeurs, ce qu'ils font d'ailleurs d'entrée, en optant pour une telle approche ! Et par la suite, avec de nombreux écrits.
- Il y a de fortes réactions au changement, réactions qui sont à la mesure des menaces ressenties.

Il reste que ce type d'approche soulève deux difficultés majeures :

- La participation requise pose le problème de l'attribution des rôles à chaque instance et à chaque groupe d'acteurs.

Quel degré de participation faut-il attribuer aux participants ? Quel équilibre faut-il trouver entre implication, autonomie et contrôle ?

L'évaluation du curriculum de français donne une réponse possible. On voit que la réponse est à chercher dans un contexte donné !

Cette question montre néanmoins que la distribution des rôles doit être claire, au départ. Si cette attribution n'est pas claire, elle peut générer conflits et insatisfaction.

• La seconde difficulté concerne la position des chercheurs et le problème de l'objectivité : la subjectivité fait-elle perdre la crédibilité des recherches ?

Cette question est débattue avec Michael Huberman dans son article "S'évaluer pour s'illusionner".

LE PROBLÈME DE L'OBJECTIVITÉ

Dans son analyse critique des modèles présentés ci-dessus, notamment de l'évaluation interactive et de la recherche-action, Michael Huberman relève le manque d'objectivité des informations recueillies par les chercheurs dans ce type d'approche et se pose la question suivante :

- Quelles sont les sources principales de biais dans le modèle d'évaluation adaptatif ?

Pour lui, les biais les plus sérieux relèvent globalement de la manière dont sont récoltées, traitées et utilisées les informations évaluatives.

De plus, en se joignant personnellement ou institutionnellement au praticien sur le terrain, le chercheur adopte les mêmes biais et les mêmes pratiques évaluatives que celui-ci.

• En fait, toute sa démonstration repose sur ce "clonage" : en s'impliquant dans la recherche, en la comprenant de l'intérieur, le chercheur devient comme le praticien. Il est soumis aux mêmes biais que lui lorsqu'il traite de l'information.

Une fois l'analogie établie, entre le chercheur impliqué et le praticien en action, Michael Huberman cite une série impressionnante de recherches en psychologie sociale expérimentale qui montrent non seulement comment l'individu récolte, traite et utilise l'information à sa disposition, mais surtout comment il fausse les données qu'il recueille.

Par quels mécanismes cela arrive-t-il ?

En résumé, on relèvera les résultats suivants :

- l'individu tient à ses opinions : devant des opinions contraires, il utilise sélectivement l'information et utilise différentes stratégies lui permettant de retenir les

opinions qui soutiennent les siennes et d'écarter celles qui s'y opposent;

- enfin, il traite l'information de manière à renforcer ses idées préconçues.

Huberman mentionne qu'en conduisant des études de ce type, les chercheurs en psychologie sociale cognitive tentaient «de comprendre les mécanismes que nous utilisons pour "raisonner" sur les faits sociaux et, en particulier, de déterminer en quoi le raisonnement de l'individu "moyen" se distingue des procédés utilisés dans la recherche expérimentale classique.» Très brièvement les commentaires essentiels sont les suivants:

"Nos raisonnements dépendent en grande partie de la stabilité ou de la cohérence de notre représentation (perceptions, attributions, attitudes) de l'univers physique et social. Nous sommes même prêts à biaiser les informations traitées afin de maintenir cette cohérence, d'où les phénomènes classiques de préjugé et autres distorsions cognitives."

- En ce qui concerne les cadres et tous ceux qui doivent prendre des décisions au sein d'une organisation, il en serait de même.

Ceux-ci cherchent surtout les informations qui:

- a) réduisent l'incertitude des choix à faire et
- b) vont dans le sens des résultats souhaités au départ.

Au sein des organisations, il y a aussi tendance à chercher de manière sélective les données appuyant son point de vue ou ses préférences, et il y a également tendance à interpréter sélectivement ces données dans le sens voulu.

Les biais se manifestent donc à trois moments:

- **lors de l'acquisition de l'information** : les études montrent que les sources facilement accessibles (orales) et affectivement connotées (celles d'un ami proche par exemple) ont plus d'influence sur nous que les sources de qualité contrôlée provenant des unités de recherches.
- **lors de sa communication** : les recherches font état de phénomènes de distorsion de l'information (bouche à oreille), de dégradation du message (par l'utilisation de registres langagiers différents, par exemple), de rétention ou de gonflement de l'information, etc.
- **lors de son traitement** : elles montrent également les limites cognitives des individus : ils ne peuvent traiter qu'un nombre limité d'informations (sept \pm 2); on

apprend aussi que deux experts en face de la même décision à prendre utilisent des informations limitées différentes !

En résumé, les conclusions des travaux empiriques portant sur l'acquisition d'informations par les gestionnaires et par les décideurs recourent celles que nous avons analysées plus haut, à savoir:

- tendance à la rétention, à la pondération et à la ré-interprétation sélectives des informations;
- biais auto-justificatifs : ils consistent à percevoir trop favorablement la réalité, notamment en ignorant les informations défavorables. Il y a aussi des surinvestissements dans les projets qui échouent;
- rétention particulière des données concrètes, dramatiques, immédiates.

QUE FAIRE?

Pour surmonter ces difficultés, Huberman propose quatre techniques permettant de retrouver une certaine rigueur, à partir du moment où l'on décide d'abandonner la rigueur expérimentale ou corrélationnelle :

1. Décentration systématique

L'idée est de confronter sa vision de chercheur à celles des autres, notamment d'autres chercheurs.

2. "Triangulation"

L'idée de base est de confronter les résultats: pour tout résultat d'importance, il devrait y avoir plusieurs constats indépendants et concordants, même si chaque constat est imparfait.

3. Prédictions et répliques

Si l'analyse pratiquée tient debout, on devrait pouvoir anticiper ce qui se produira dans les mois à venir. De même, si l'analyse joue, on devrait pouvoir la refaire dans un autre coin de son échantillon que l'on n'a pas étudié en détail ou, le cas échéant, on devrait pouvoir constater son absence dans une situation contrastée.

4. Documentation

Comme dans les recherches classiques dotées d'une convention dans la publication des résultats, l'idée est de proposer au chercheur de produire un rapport qui permette de déterminer le bien-fondé des résultats ou des conclusions. Le lecteur doit pour le moins savoir ce qui a été utilisé pour récolter, mettre en forme, analyser et interpréter les informations. D'où la nécessité, tout simplement, de le lui dire.

EN CONCLUSION

Toutes les critiques et tous les conseils de Michael Huberman s'appuient sur le fait que le chercheur, en s'impliquant lui-même dans la recherche, ne peut plus se distinguer des praticiens avec lesquels il travaille.

- Au niveau des représentations et des enjeux symboliques ou réels de pouvoir, cela semble douteux.

- En ce qui concerne les recherches traitant des biais de l'information, toutes arrêtent leurs analyses et leurs conclusions à un premier niveau sémantique des messages.

Que faire alors des non-dits, des informations cachées, et de tous les niveaux symboliques et représentatifs se servant du langage simplement comme support à une communication qui ne se révèle qu'à l'interprétation.

Il semble que l'utilité du chercheur impliqué réside dans le fait qu'il a justement pour tâche de mettre en lumière tous ces niveaux de la communication humaine, afin de bien comprendre et de bien faire comprendre la situation !

- Enfin, en ce qui concerne le problème général de l'objectivité, les dangers sont réels. Y ajouter une panoplie de garde-fous techniques, est-ce suffisant ?

D'autres chercheurs ont proposé l'accord intersubjectif et la négociation afin de garantir une certaine pertinence à la fois dans les décisions qui sont prises et dans les interprétations qui sont données.

BIBLIOGRAPHIE

- J. Cardinet, Les modèles d'évaluation applicables aux innovations pédagogiques, Contributions de Jean Cardinet au Séminaire international sur les stratégies de l'innovation pédagogique dans l'enseignement supérieur organisé par l'AUPELF, octobre 1976, Neuchâtel : IRDP, Service de la Recherche, 1976
- J. Cardinet, L'évaluation du nouvel enseignement de la mathématique en première année primaire : présentation de l'expérience romande, Neuchâtel : IRDP/R 77.17, Août 1977.
- J. Cardinet & J. Weiss, L'observation interactive., au confluent de la formation et de la recherche, Neuchâtel : IRDP/R 78.16, 1978
- J. Cardinet, Savoirs généraux et savoirs locaux, Neuchâtel : IRDP/R, 1980.
- J. Cardinet (1990), L'histoire du RIRELF, in Le réseau international de recherche en évaluation de langue française (R.I.R.E.L.F.) : Une expérience de cinq années (1985-1990), France : COREP, pp. 108 - 125
- M. Huberman, S'évaluer pour s'illusionner ? Promesses et écueils de l'évaluation "adaptative/interactive" des innovations scolaires, Neuchâtel : IRDP/R 83.08, Cahier du GCR/SSRE no 9, 1983
- M. Jeannet, Aspects psychosociologiques de l'évaluation : Exposé présenté par l'auteur à la séance du Groupe des chercheurs romands (GCR) du 25 janvier 1977, à Lausanne, Neuchâtel : IRDP/R 77.14, 1977
- G. de Landsheere, Dictionnaire de l'évaluation et de la recherche en éducation, Paris : PUF, 1992 (2ème édition).

CH.8B LES ENQUÊTES D'ÉVALUATION DES SYSTÈMES SCOLAIRES EUROPÉENS

Origine

La pédagogie scientifique centenaire

Les statistiques sur le nombre d'élèves, de diplômés, etc. ont accompagné dès le début le développement des systèmes scolaires. Elles se sont surtout multipliées dans tous les pays depuis l'arrivée de l'informatique, mais sans coordination, pour répondre à des besoins uniquement administratifs et nationaux.

Les enquêtes pour mesurer les performances scolaires ont commencé à la fin du siècle dernier, avant même le développement des tests psychologiques, mais ont été conduites par d'autres instances, davantage soucieuses de pédagogie, par des inspecteurs notamment.

Il a fallu que soit créée la pédagogie expérimentale, surtout après la seconde guerre mondiale, pour que la mise en relation des deux types d'enquêtes devienne un moyen de recherche opérationnelle.

Mais l'interprétation de ces données était rendue difficile par l'absence de points de comparaison avec d'autres pays.

La base des travaux de l'OCDE

L'existence de ces données statistiques et l'exemple des premières recherches de l'IEA (voir plus bas) ont suscité le besoin de mesures objectives pour décrire le fonctionnement des systèmes scolaires, sur le modèle des statistiques économiques internationales, déjà mises au point par l'OCDE.

Les enquêtes françaises

Le rôle de la DEP

Les premières enquêtes françaises ont été réalisées dans les années 70. Pour assurer leur continuité et les développer, la Direction de l'Évaluation et de la Prospective a été créée en 1987; c'est elle qui organise actuellement pour le Ministère de l'Éducation, (mais en s'entourant des avis des divers acteurs sociaux présents dans l'école), les travaux d'évaluation dans les trois directions suivantes:

Les acquis des élèves

Des cohortes de plusieurs milliers d'élèves sont suivies de façon longitudinale à travers toute leur scolarité, sur les plans cognitifs, (développement des connaissances et aptitudes), mais aussi des attitudes, de la culture générale, de l'adaptation à l'école, de l'orientation professionnelle, etc.

Les facteurs d'apprentissage

Ces données peuvent alors être mises en rapport avec les diverses facettes du contexte familial et scolaire des élèves, pour étudier les conditions sociales qui ont une influence sur les performances scolaires. On peut ainsi apprécier par ex. l'effet de mesures d'appui ou de sélection, ou bien celui de structures scolaires expérimentales, sur la réussite ultérieure des élèves de chaque niveau socioculturel, ce qui peut guider les décisions administratives ou politiques du Ministère.

Les moyens d'évaluation

Des épreuves standardisées sont distribuées aux enseignants pour qu'ils puissent situer en début d'année le niveau de leur classe et les difficultés particulières de certains élèves. L'intention est de guider les mesures correctives individualisées des maîtres dans leur classe.

Les enquêtes anglaises*Le rôle de l'APU (Assessment of Performance Unit)*

Le Ministère de l'Éducation et de la Science a créé cette organisation dès 1975, au départ pour ses besoins.

Très vite, cependant, les résultats ont été diffusés, mais en respectant l'anonymat des élèves et des écoles.

Les "surveys" de l'APU ont permis de situer le niveau de performance des élèves anglais en langue maternelle, mathématique, sciences, développement personnel et social, développement esthétique, et développement physique.

Le rôle du SEAT (School Evaluation and Assessment Council)

La loi de réforme de l'éducation de 1988 a introduit un curriculum national, avec des objectifs opérationnalisés, et une évaluation nationale pour contrôler dans quelle mesure ces objectifs éducatifs prescrits sont atteints par chaque école.

Dix disciplines sont contrôlées à quatre moments dans les études. Les résultats sont rendus publics, par école et par classe, et le financement des écoles peut être affecté par ces résultats (principe de la libre concurrence).

Le niveau de chaque élève est situé dans chaque discipline sur une échelle critériée à dix échelons, ce qui devrait permettre aux enseignants de partir avec chaque élève du niveau où il se trouve et d'individualiser l'enseignement.

Les indicateurs de performance des LEA (Local Educational Authorities)

Pour chaque district scolaire, des mesures vont être publiées périodiquement sur le taux d'encadrement des élèves, les diplômes des enseignants, l'utilisation du temps, les devoirs donnés à domicile, l'absentéisme, etc.

Les enquêtes suédoises*Le système de modération*

Des épreuves sont préparées sur un plan national pour permettre à chaque enseignant de mieux situer ses exigences en sachant où ses élèves se situent par rapport à ceux de l'ensemble du pays.

L'évaluation nationale

Les connaissances et les compétences des élèves sont évaluées dans chaque discipline, tous les trois ans, pour savoir si les objectifs pédagogiques sont atteints, école par école et pour l'ensemble du pays.

Une évaluation qualitative de chaque école est aussi conduite chaque année

Les enquêtes IEA

Leur origine

Ce sont des chercheurs en éducation qui ont effectué une première recherche en 1959-62 sur les connaissances scolaires des enfants de 13 ans dans une dizaine de pays.

L'enquête sur la mathématique

Ce projet pilote a été suivi d'une grande enquête sur l'enseignement des mathématiques, qui a duré de 1961 à 1967. C'est cette enquête qui révéla que le Japon avait les résultats scolaires les meilleurs du monde. Le choc que provoquèrent ses résultats dans les pays occidentaux fit beaucoup pour stimuler la réforme de l'enseignement de la mathématique, en même temps que le développement des méthodes de l'éducation comparée et la remise en question des pratiques d'évaluation scolaire.

Le survey des six disciplines

22 pays y ont participé, de 1966 à 1973. Cette fois, c'est surtout le contexte qui a été étudié, au moyen de plusieurs centaines de variables dont on a cherché à étudier l'effet sur le rendement.

La 2ème enquête sur l'enseignement des mathématiques

En plus d'un bilan des performances obtenues, on a cherché à mesurer l'augmentation des connaissances au cours d'une année scolaire, et à en rendre compte, au niveau individuel et à celui des classes. (1980-86)

Les recherches récentes

- 2ème enquête sur les sciences: 1984
- Enquête sur la compréhension de la lecture (1990-93)
(La Suisse y a participé, bien que tous les cantons n'aient pas accepté de se soumettre à l'étude.)
- Etude sur l'utilisation pédagogique des ordinateurs (1990)
(Là encore certains cantons n'ont pas participé.)
- Etude sur l'éducation préscolaire (1986-1996).

Influence de ces enquêtes

L'IEA a joué un grand rôle en posant de bonnes questions et en mettant en lumière parfois les insuffisances des moyens statistiques à disposition pour y répondre.

L'OCDE reprend finalement à son compte une partie de ses objectifs, en essayant de jeter les bases qui permettront à l'avenir des comparaisons internationales significatives.

Les gouvernements ont cependant très peu tenu compte jusqu'ici de l'expérience des autres pays. (Même les cantons suisses refusent d'admettre comme valable chez eux une expérience menée dans un canton voisin!)

La lourdeur et la lenteur de ses travaux, liées à la difficulté d'obtenir des financements pour une organisation de recherche indépendante, ont récemment suscité le développement d'une organisation concurrente, l'IAEP.

Les enquêtes de l'IAEP (International Assessment of Educational Progress)

Le rôle de l'ETS (Educational Testing Service)

Cette organisation s'est spécialisée aux Etats-Unis dans la construction de tests et leur dépouillement par ordinateur. Ses chercheurs ont développé la Théorie de réponse aux items, modèle statistique qui permet de situer sur une même échelle des réponses à des questions différentes. Ils peuvent ainsi effectuer des recherches qui n'étaient pas possibles avec des épreuves classiques, et aussi traiter les résultats plus rapidement. Ils ont obtenu du gouvernement américain la charge de suivre les résultats scolaires état par état aux USA, et ils se sont lancés en 1987 dans une démarche analogue au niveau international.

L'enquête I : Mathématiques - Sciences (1988)

Les connaissances des élèves de 13 ans ont été étudiées dans cinq pays, à partir de questions déjà élaborées aux Etats-Unis, mais adaptées aux circonstances de chaque région.

L'enquête II Mathématiques - Sciences - Géographie (1990-91)

L'étude précédente a été reprise dans 28 pays et à 9 et 13 ans. La Suisse y a participé et y apparaît en bonne place, bien que la représentativité des échantillons soit toujours un problème dans un pays comme le nôtre.

Résultats de ces enquêtes

Montrer la nécessité de mesures plus comparables

Si les études purement nationales sont d'emblée privées de toute valeur générale, les enquêtes internationales sont encore très loin de répondre aux exigences de comparabilité: même si l'on trouve des questions qui peuvent être traitées par tous les élèves, ce sous-ensemble commun n'est pas relié de la même manière à l'ensemble du curriculum de chaque pays.

On retrouve le problème des examens nationaux, où brillent les élèves qui ont eu la chance d'être formés dans la même perspective que celle de l'examineur; d'autres examens auraient favorisé d'autres classes.

Mais révéler l'arbitraire des concours

L'erreur philosophique de base est sans doute de vouloir mettre les systèmes scolaires en compétition et de ramener ainsi l'excellence à une dimension unique, alors qu'elle est multidimensionnelle et ne permet pas le classement de premier à dernier. Tout concours repose sur le choix arbitraire d'une échelle de valeur particulière.

Montrer la spécificité de chaque situation

Cette réflexion critique ne met pas en cause l'intérêt de comparaisons qui permettent de prendre conscience des choix de valeurs, trop souvent inconscients, des systèmes éducatifs. Pour développer une identité, il faut se choisir différent des autres, et pas seulement supérieur ou inférieur.

Importance relative (en pourcentages) des différents contenus mathématiques dans l'étude de l'IAEP de 1991, comparée aux objectifs déclarés, dans deux pays

<i>Contenus mathématiques</i>	<i>Questionnaire de l'IAEP</i>	<i>Objectifs Israël</i>	<i>Objectifs Suisse</i>
Nombres et opérations	30	10	50
Mesures	15	10	10
Géométrie	20	25	20
Statistiques et probabilités	15	5	10
Algèbre et fonctions	20	50	10
Total	100	100	100

Israël et la Suisse (14cantons) sont deux des vingt pays qui ont participé à l'étude

Source: IAEP, The 1991 IAEP assessment: Objectives for mathematics, science and geography, p. 17, Princeton, Center for the Assessment of Educational Progress, Educational Testing Service, 1991.

Veuillez photocopier juste l'encadré
et rajouter aussi 5 pages de PB (revenir)

Encadré 4.7

Facteurs influant sur la maîtrise de la lecture chez les écoliers de 32 pays : conclusions de l'étude de l'IEA 1990-1991

Différences entre pays à score élevé et pays à score faible.

Les facteurs qui, de façon constante, sont présents dans les pays à score élevé, par opposition aux pays à score faible, sont les suivants : importantes bibliothèques scolaires, importantes bibliothèques de classe, emprunts d'ouvrages réguliers, fréquentes lectures silencieuses en classe, lecture fréquente d'histoires à haute voix par les enseignants, emploi du temps faisant une plus large place à l'enseignement de la langue. Plusieurs pays à score faible ont, d'après les renseignements qu'ils ont fournis, une expérience très limitée des tests de lecture de type formel mais, au-delà d'un certain seuil, il est apparu que ce n'était pas là un facteur pertinent de différenciation entre pays à score élevé et pays à score faible. Chez les enfants de 14 ans, on relève aussi d'autres facteurs de différenciation : travail à la maison plus généralisé, école offrant plus de moyens d'alphabétisation, enseignement plus individualisé, moins d'enseignants n'ayant pas la langue enseignée pour langue maternelle, année scolaire plus courte, plus forte proportion d'enseignants de sexe féminin.

Facteurs de différenciation d'importance secondaire.

Chez les écoliers de 9 ans, il n'est pas apparu que les niveaux d'alphabétisme soient sensiblement supérieurs dans les pays où l'on note des taux élevés de préscolarisation, des classes généralement moins chargées, de nombreuses classes regroupant plusieurs niveaux d'études, des années scolaires plus longues, ou un suivi de leur classe par les enseignants pendant plusieurs années successives. S'il s'est avéré que ces facteurs sont généralement présents dans les pays à score élevé, les données dont on dispose laissent à penser qu'ils s'expliquent peut-être simplement par une relative richesse et par certaines caractéristiques extrascolaires de la collectivité.

Différences selon le sexe. Les filles ont obtenu de meilleurs scores que les garçons dans tous les pays pour ce qui est de la population A (9 ans), et dans la plupart des pays pour ce qui est de la population B (14 ans), la différence moyenne en faveur des filles tombant de 12 points à 7 à l'âge de 14 ans. [...]

Différences selon la langue. Dans tous les pays, les enfants qui parlent à la maison une autre langue qu'à l'école ont, pour les deux âges considérés, un moins bon niveau d'alphabétisation. [...]

Différences entre zones urbaines et zones rurales. Les enfants des zones urbaines obtiennent de meilleurs résultats que les enfants des zones rurales dans la plupart des systèmes éducatifs. Dans quelques pays très développés, cependant, les élèves ruraux témoignent d'un niveau d'alphabétisation aussi élevé que leurs camarades des villes, voire supérieur.

Importance des livres. L'accès aux livres est un facteur clé de la maîtrise de la lecture. En règle générale, les pays qui ont les meilleurs scores sont ceux où les enfants se procurent le plus facilement des livres, chez eux, dans des bibliothèques publiques et des librairies proches de leur domicile, et à l'école.

Liens avec la télévision. En dehors de l'école, la télévision occupe une bonne partie du temps libre des élèves. Dans plusieurs pays, nombreux sont les enfants qui passent plus de cinq heures par jour devant le poste. Ceux qui regardent souvent la télévision tendent, en règle générale, à avoir de moins bons résultats que les autres. Cependant, dans quelques pays, ce sont les enfants qui regardent la télévision de trois à quatre heures par jour qui lisent le mieux.

Source. Warwick B. Elley, *How in the world do students read?*, p. xii-xiii, La Haye, Association internationale pour l'évaluation du rendement scolaire (IEA), 1992.

l'étude quant aux raisons qui expliquent les différences entre pays pour ce qui est des performances d'ensemble. Les observations faites sont en elles-mêmes d'un intérêt évident mais elles ne fournissent aucune indication sur l'efficacité relative des différents systèmes éducatifs. Dans quelle mesure l'élève savait-il déjà lire à son entrée à l'école? En l'absence d'information sur ce point (l'étude n'en donne pas) il est impossible de dire si la lecture est mieux enseignée dans un pays que dans un autre — en d'autres termes, si la «va-

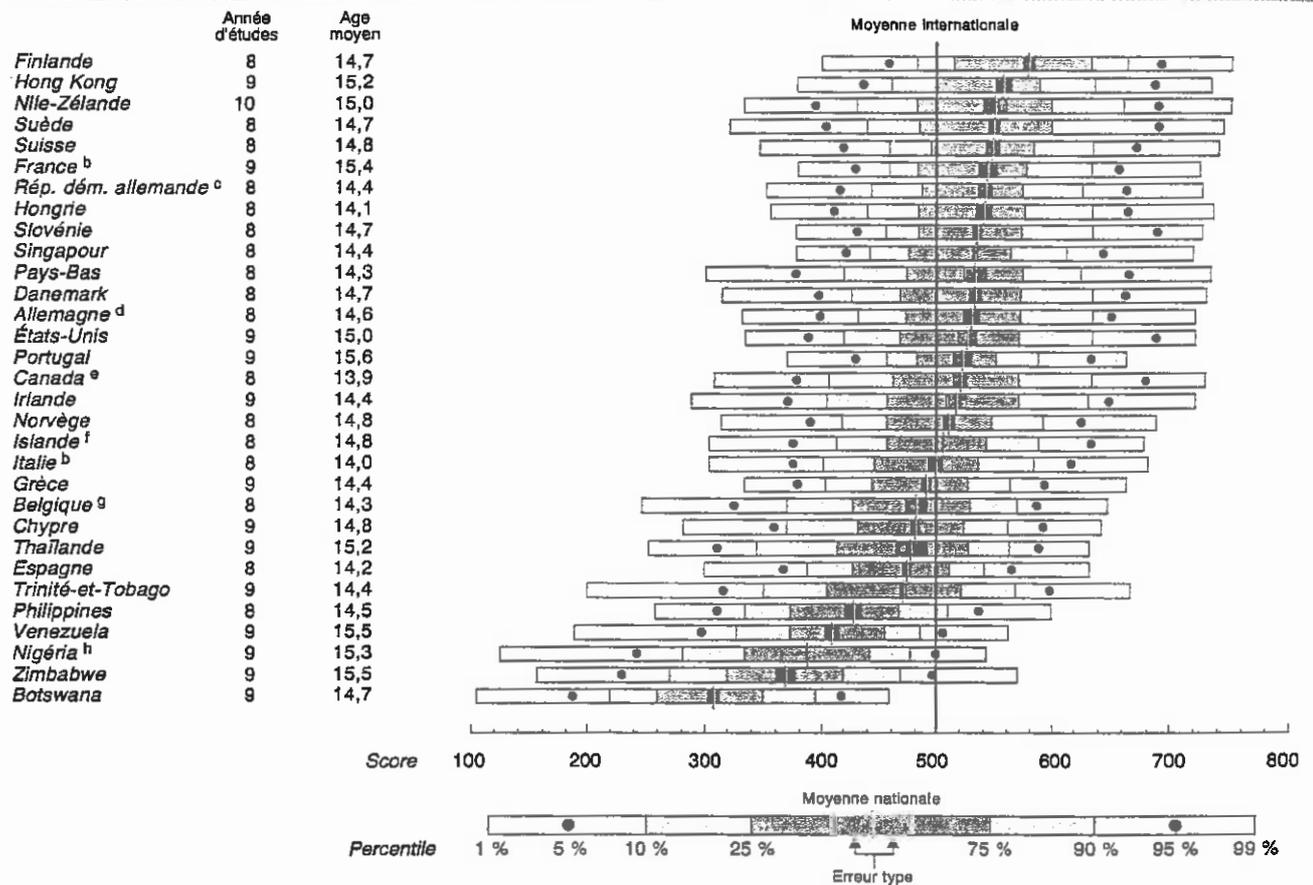
leur ajoutée», dans ce cas, par le système d'enseignement est supérieure dans tel pays ou dans tel autre.

Cette question de la «valeur ajoutée» est particulièrement pertinente dans le cas des pays en développement, qui se classent presque invariablement en queue lors de ces «olympiades de la connaissance». Comme on l'a déjà noté plus haut, la langue d'enseignement, dans un nombre sans doute majoritaire de pays de l'Afrique subsaharienne par exemple, n'est pas, pour beaucoup d'élèves (voire souvent la plupart) leur

Essayer d'agrandir et d'en faire un transparent (entre les 2 feuilles) y joindre une autre photocopie meilleure pour avoir l'original

Figure 4.3

Efficacité de l'apprentissage ou « opportunités d'apprentissage » de la lecture ? Scores^a des élèves de 14 ans au test de lecture de l'IEA dans 31 pays, 1990-1991



- a. Scores de lecture des « documents » (voir le corps du texte).
 b. Établissements de l'État seulement.
 c. Ex-République démocratique allemande.
 d. République fédérale d'Allemagne d'avant la réunification.
 e. Colombie britannique seulement.
 f. Toute la population étant couverte par l'étude, il n'y a pas à calculer d'erreur type.
 g. Population francophone seulement.
 h. Les données sont insuffisantes pour permettre de calculer l'effet du plan expérimental.

Source. Warwick B. Elley, *How in the world do students read ?*, p. 30, La Haye, Association internationale pour l'évaluation du rendement scolaire, 1992.

par exemple le logement et la famille (pour les élèves), la classe et les méthodes pédagogiques (pour les enseignants) et la situation dans l'établissement (pour les chefs d'établissement). Trente-deux pays au total ont participé à l'étude. Chaque évaluation par questionnaire couvrait trois grands domaines ou types de textes de lecture :

Prose narrative : textes suivis dont l'auteur se propose de raconter une histoire, vraie ou inventée.

Prose descriptive : textes suivis visant à décrire, expli-

quer ou communiquer d'une autre manière au lecteur un ensemble d'informations factuelles ou d'opinions (par exemple, brèves lettres à la famille, descriptions d'animaux, etc.).

Documents : présentation d'une information structurée sous forme de graphiques, tableaux, cartes, courbes, instructions ou modes d'emploi.

La figure 4.3 présente les résultats globaux (scores) du test administré aux élèves de 14 ans sur les *documents*, et l'encadré 4.7, les conclusions générales de

Importance relative (en pourcentages) des différents contenus mathématiques dans l'étude de l'IAEP de 1991, comparée aux objectifs déclarés, dans deux pays

<i>Contenus mathématiques</i>	<i>Questionnaire de l'IAEP</i>	<i>Objectifs Israël</i>	<i>Objectifs Suisse</i>
Nombres et opérations	30	10	50
Mesures	15	10	10
Géométrie	20	25	20
Statistiques et probabilités	15	5	10
Algèbre et fonctions	20	50	10
Total	100	100	100

CH. 10. B. LA NOTION DE RÉFÉRENTIEL

1) Nécessité d'un référentiel

En accord avec la plupart des auteurs, il n'est en effet pas envisagé d'évaluation sans recours implicite ou explicite à un système de références.

Par rapport à quoi évalue-t-on ? La réponse est partiellement donnée : des références ! Mais que l'on évoque un système ou un cadre de références, un ou des référents, un ou des référentiels, on ne dit rien sur la nature de ces références.

En invoquant des références, on introduit aussi le problème de la légitimité de leur choix. Comment les justifier ?

Et qui aurait compétence pour les élaborer ?

- les décideurs, propriétaires des orientations et des priorités;
- les chercheurs, détenteurs d'instruments d'observation et d'analyse;
- les experts et les contrôleurs, gardiens des normes;
- les acteurs participant à l'action éducative.

L'évocation même de toutes ces questions suggère que la définition ou la recherche du référentiel se pose dans une perspective plurielle.

En posant que peut-être tous les partenaires seraient concernés (décideurs, chercheurs, experts, corps éducatif et éduqués), on entre dans les types de travaux qui suivent des démarches participatives et négociées et qui se fondent sur une épistémologie intersubjectiviste.

- Maintenant, que l'on adopte cette position épistémologique ou non, la question :

qu'est-ce qu'un cadre ou un système de références ?

reste pertinente pour l'ensemble des approches étudiées jusqu'ici.

2) Qu'est-ce qu'un cadre de références, qu'est-ce qu'un référentiel ?

Polysémie du référentiel

Le référentiel homogène

Le *référentiel* (ou le *référent*, qui serait une partie du référentiel) se conçoit comme un ensemble cohérent de normes, de règles, de valeurs qui permet une comparaison entre un objet soumis à examen et un objet "idéal" et normé (défini par des textes ou construit pour la circonstance).

Dans cette perspective, par exemple, le mètre étalon est le référent de tous les mètres utilisés; le corrigé type est celui des copies d'examen; le code de la route, est la référence du conducteur et du gendarme; les articles de la loi sur le travail est la référence légale en cette matière.

Toujours selon cette perspective, le référentiel, ensemble de référents, indique les divers points de vue à partir desquels il est nécessaire de se situer pour regarder l'objet à évaluer. Il se traduit en une série de normes précises, connues. Il s'agit d'une construction intellectuelle qui est et reste extérieure à l'évaluateur. En ce sens le référentiel doit être explicité par l'évaluateur et reconnu par l'évalué. Les écarts qu'il permet de mesurer doivent être vérifiables par les participants.

Défini ainsi, le référentiel rejoint une approche de type contrôle examinée par J. Ardoino. Il s'applique plus ou moins bien à différents champs particuliers du domaine économique et social qui ont recours pour l'essentiel au

normatif. Dans ces cas, le contrôleur est interchangeable; quant au référent, il est clair, stabilisé, homogène.

Le référentiel hétérogène

En matière d'éducation et de formation, lorsqu'il s'agit d'évaluer des établissements scolaires, la détermination des référentiels est moins aisée.

Les sources et les origines des référentiels sont multiples et elles posent la question du choix. On peut distinguer, à titre d'exemple et sans souci de classification hiérarchique ou d'exhaustivité :

- une origine officielle (externe) : les lois, les règles juridiques, les circulaires, etc.;
- une origine interne : tout ce qui fait exister et fonctionner l'institution à l'intérieur (notes de services, emploi du temps, répartition en classes, décisions du conseil d'administration);
- une origine externe : tout ce qui fait exister et fonctionner l'institution à l'extérieur (environnement politique et social, conventions passées avec d'autres instances) et données externes (statistiques nationales);
- une origine venant de l'expertise : évoquer un cadre ou un système de références peut aussi souvent vouloir dire désigner un champ scientifique par rapport auquel l'explication d'un phénomène va être fournie : par exemple, on va expliquer le fonctionnement de l'établissement à partir des savoirs et des méthodes de la sociologie des organisations ou de l'analyse de systèmes, etc. Ces champs de savoirs sont alors pris comme systèmes de références déjà constitués et capables d'apporter, chacun, un sens possible à certains résultats. Dans cette perspective, le référentiel constitue un type de lecture, une grille d'analyse.

Cependant, cette diversité dans la composition des sources du référentiel montre que si le cadre de références peut se définir en un ensemble de normes, de règles et de valeurs, celles-ci peuvent présenter des natures fort diverses :

- normes juridiques (lois, règlements);
- normes prospectives (objectifs à atteindre, prévisions);
- normes spatiales : comparaisons sur le plan national, international;
- normes temporelles et contextuelles : prise en compte de l'histoire, de l'évolution de l'établissement;
- normes culturelles : qui englobent à la fois des normes spatiales et temporelles.

On remarque ainsi que plus on prend en compte des normes de nature différente, plus on s'éloigne du contrôle proprement dit. Les référents sont multiples, peu stables et hétérogènes; il est en outre difficile, voire impossible de substituer un évaluateur à un autre.

Les difficultés

Quelles sont alors les principales conséquences de cette multiplicité des référentiels ? En bref :

- Il n'y a pas d'harmonie préétablie entre les divers référentiels. Certains sont en conflit direct. Au départ, la cohérence n'est pas donnée, mais elle est à construire.

- De cette situation conflictuelle, il ressort que toute tentative d'articuler les divers types de référentiels entre eux se heurte souvent à la divergence, à la contradiction ou à l'incompatibilité.
- En supposant que les référentiels soient harmonisés, coordonnés et définitivement établis, le travail sera à recommencer à brève échéance, car leurs durées sont variables et limitées. En effet, si les normes changent, les référentiels qui s'y rapportent changent aussi.
- La construction et la mise à jour perpétuelle du ou des référentiels sont donc inévitables.
- Une dernière difficulté est bien sûr de sélectionner ce qui est pertinent pour la fonction étudiée, dans l'établissement donné et pour une demande précisée.

Les dangers

On voit que le référentiel peut être différemment utilisé. Il peut se rapporter à une panoplie de normes, on l'a vu, mais aussi aux savoirs de diverses disciplines scientifiques et encore, dans d'autres cas, à des catalogues d'objectifs hiérarchisés. La liste n'est sûrement pas exhaustive.

G. Figari note que l'on peut ainsi glisser de l'un à l'autre, de l'objectif à la norme par exemple, sans s'en rendre compte.

Il fait également remarquer que les référentiels, une fois établis et devenus normatifs, sont amenés à leur tour, pour faire preuve de leur légitimité, à citer leurs références, à dire par rapport à quoi ils fonctionnent : il y a donc recherche sans fin du référentiel du référentiel.

3) Détermination du référentiel ou des critères¹ de l'évaluation.

Selon Nadeau (1988) :

L'identification des critères d'évaluation représente l'une des tâches les plus difficiles à accomplir.

Selon Nadeau, les auteurs s'entendent pour dire que les critères doivent être déterminés selon :

- le contexte particulier dans lequel se situe l'objet
- ainsi que selon la fonction de son évaluation.

Nadeau résume les avis de la manière suivante :

- Certains suggèrent que l'atteinte des objectifs soit le critère pour juger de la valeur de l'objet. (Tyler, Provus, années 50-70's)
- D'autres contournent le problème du choix des critères en l'ignorant ou encore en insistant sur le fait que l'évaluateur n'est là que pour fournir de l'information au preneur de décision. Le problème du choix est alors laissé au décideur. (Cronbach, Stufflebeam, années 60's)
- D'autres considèrent plutôt les besoins actuels et potentiels des clients. (Joint Committee, Scriven, années 70 et 80's)
- La prise en compte des valeurs idéales et sociales est encore une autre possibilité. (Guba, années 80's)

¹ G. Figari définit le critère comme le principe qui permet de juger, de distinguer le vrai du faux, d'estimer, d'apprécier. Sa première caractéristique est l'abstraction (par exemple, la rapidité, la clarté, l'évidence, l'ordre, etc.). Sa seconde caractéristique est la discrimination : 1) entre les objets possédant ou ne possédant pas le critère retenu; et 2) à l'intérieur des objets possédant le critère : distinction qualitative. Quant aux indices et aux indicateurs, ils viendraient à l'étape suivante et sont de nature opératoire.

Selon Lecointe et Rebinguet (1990) :

Même si ces deux auteurs consacrent leur ouvrage à l'audit (procédé dans lequel ils incluent l'évaluation, ce qui n'est pas toujours le cas selon les auteurs), on peut citer, en transposant, ce qu'ils disent au sujet de la détermination du référentiel de l'audit :

Ce qui guide le choix dans les sources et dans la panoplie des niveaux de référence, c'est l'objectif de l'évaluation, objectif qui détermine un type de référentiel.

En gros et pour un établissement scolaire, on aura trois types d'évaluation et donc trois types de référentiel :

- **Évaluation de conformité :**

La question posée est par exemple : comment innover tout en restant en conformité ?

Le référentiel de la conformité, ce seront les textes, le règlement.

- **Évaluation d'efficacité :**

La question est : par exemple, comment conduire 74% de nos élèves au bac ?

Le référentiel de conformité est second (mais pas abandonné !); le référentiel principal est un référentiel d'objectifs (l'ouverture de sections, la réduction des redoublements, des normes de soutien et d'amélioration de la productivité de l'enseignement) et sa compatibilité avec le référentiel de conformité.

- **Évaluation stratégique :**

On est conforme, mais on n'a pas d'objectifs propres : comment s'en donner ? Comment faire ?

Le référentiel est la cible stratégique; il est à construire, à partir de l'existant et par projections successives (deux, trois, cinq ans...); avec les acteurs de l'établissement et pas à leur place.

Les deux auteurs concluent : à la lecture de ces exemples, il n'y a jamais un seul type d'évaluation (d'audit pour eux) mis en oeuvre de façon chimiquement pure. Par conséquent, les référentiels correspondants sont des dominantes qui vont se traduire dans la construction du référentiel particulier choisie pour l'évaluation (l'audit).

Selon G. Figari (1994) :

Selon cet auteur, c'est la procédure elle-même de recherche du référentiel, c'est la manière dont on établit le référentiel qui comporte de l'intérêt et non le produit fini qui peut être légitimement remis en cause par l'utilisateur suivant.

Un exemple tiré de son ouvrage est présenté par Jean Cardinet.

Bibliographie

Ardoino, J. & Berger, G. (1989). D'une évaluation en miettes à une évaluation en actes. France : ANDSHA/Matrice.

Cardinet, J. (1987). Évaluation externe, interne ou négociée ? Neuchâtel : IRDP/R 87.114.

De Landsheere, G. (1992). Dictionnaire de l'évaluation et de la recherche en éducation. Paris : PUF, (2ème édition).

Figari, G. (1994). Évaluer : quel référentiel ? Bruxelles : De Boeck Université.

- Le Boterf, G., Dupouey, P. & Viallet, F. (1985). L'audit de la formation professionnelle. Paris : Les Éditions d'Organisation.
- Lecoite, M. & Rebinguet, M. (1990). L'audit de l'établissement scolaire. Paris : Les Éditions d'Organisation.
- Nadeau, M.-A. (1988). L'évaluation de programme : Théorie et pratique. Québec : Les Presses de l'Université Laval, pp. 44-51.

CH. 13 LE MODÈLE DU MICHIGAN

Un "Guide pratique pour l'évaluation de programmes" a été publié par le Centre d'Évaluation de la Western Michigan University (auteurs: Robert Brinkerhoff, Dale Brethower, Terry Hluchyj, et Jeri Ridings Nowakowski. Editeur: Kluwer-Nijhoff Publishing, Boston, 1983). Il va servir à regrouper en sept étapes les différentes questions à traiter pour évaluer un curriculum ou le fonctionnement d'une institution. (On parlera de "programme".)

1. Poser le problème

1.1. *Que veut-on évaluer?*

Bien qu'essentielle, cette question est particulièrement délicate et il faut y revenir cycliquement après chacun des points envisagés sous 1. par lesquels on cherche à mieux décrire la situation pour mieux poser le problème.

Traits à inclure dans la description:

- les acteurs (décideurs, pourvoyeurs de fonds, administrateurs, personnel enseignant, enseignés, familles, syndicats, associations, opposants).
- les buts du programme (besoins connus, objectifs annoncés, conflits d'objectifs, buts des divers acteurs, évolution des besoins et des objectifs).
- les composantes du programme (décomposition du programme, dans le temps et dans le contenu, relations entre ces parties, problèmes d'équilibre).
- les activités déclenchées par le programme (enseignement, formation, entretien, administration, services extérieurs...).
- le réseau social (programme public ou privé, autres programmes similaires, activités complémentaires, sympathies et appuis extérieurs, concurrence, soutiens et opposants idéologiques et politiques).
- les ressources financières (budget, aides extérieures, installations) et humaines (personnel, bénévoles, appuis scientifiques et techniques).
- l'histoire (durée de fonctionnement, crises passées, gestion de ces crises, conséquences actuelles, perspectives d'avenir, interrogations des acteurs).
- l'environnement (quels sont les facteurs dans la communauté locale et dans la région qui peuvent influencer le fonctionnement du programme, quelles personnalités ou sociétés s'intéressent ou s'opposent au programme?)

Analyse de la situation et explicitation du problème

- La description ci-dessus suggère-t-elle des sources de difficultés probables pouvant affecter le fonctionnement du programme?
- Quelles décisions majeures sont à prendre à propos du programme?
- Quels problèmes perçoivent les différents acteurs interviewés?
- Ces acteurs ont-ils la même perception des sources de difficultés?
- Quelles solutions proposent-ils spontanément?

1.2 *Pourquoi veut-on une évaluation?*

Raisons possibles:

- Clarifier les besoins auxquels on doit répondre et les objectifs du programme (Les besoins initiaux sont-ils encore manifestes? D'autres semblent-ils plus urgents? Les priorités sont-elles à modifier dans les objectifs? Y a-t-il accord sur les objectifs prioritaires parmi les différents acteurs?)
- Repenser la façon de répondre au besoin (Faut-il résoudre des conflits entre les acteurs du programme sur la stratégie d'action? Ceci amène-t-il à comparer différents plans d'action possibles?)

Faut-il plutôt détecter les sources de dysfonctionnement? Faut-il clarifier les rôles de certains acteurs?)

- Suivre la mise en application du projet d'action initial
(Le programme est-il appliqué comme prévu? Pour quelles raisons des modifications ont-elles été apportées? Par qui? Quelles conséquences positives et négatives en découlent? Pour qui? Quels acteurs sont-ils inquiets?)
- Contrôler les résultats
(Que pensent chacun des groupes d'acteurs des résultats intermédiaires? à plus long terme? Que pensent-ils de l'investissement que ces résultats ont nécessité? L'utilité est-elle suffisante? Qualité et quantité sont-elles satisfaisantes? L'effort peut-il être maintenu ainsi à l'avenir? Que penser des effets indirects? Comment chacun voit-il l'avenir?)
- Décider de la poursuite du programme
(Faut-il abandonner le programme initial? Faut-il l'adapter? Que pensent les acteurs des directions d'amélioration qui s'imposent? Faut-il investir plus de ressources, ou économiser les moyens? Faut-il modifier les objectifs? Faut-il leur donner un poids relatif différent?)

Évaluation de ces raisons

Une évaluation n'est pas toujours souhaitable. Il vaut mieux ne pas l'entreprendre que de risquer de susciter davantage de problèmes encore.

Critères de l'acceptation:

- But clair (compréhensible par tous les groupes d'acteurs?)
- Participants informés et volontaires (le but a-t-il été explicité et les personnes concernées ont-elles été mises au courant, sinon consultées, et sont-elles d'accord?)
- Utilisation envisagée (les acteurs semblent-ils disposés à tenir compte des résultats pour adapter leur action?)
- Résultat pertinent (la question que l'on pose pourra-t-elle trouver sa réponse dans l'information que l'on veut rechercher?)
- Effets contrôlés (l'opération pourra-t-elle se dérouler jusqu'au bout sans déclencher des crises pires que les problèmes actuels?)
- Démarche compatible avec le contexte social (les groupes d'acteurs, les responsables politiques, l'opinion publique, peuvent-ils admettre l'idée d'une éventuelle remise en question du programme?)
- Travail justifié (les effets positifs attendus justifient-ils le coût, l'effort, le délai causés par une telle étude?)

Raisons de renoncer:

- La décision a déjà été prise.
- Des décisions devront être prises avant qu'on ait obtenu les résultats.
- On ne sait pas qui se sentira engagé à tenir compte des résultats.
- Certaines personnes risquent de faire les frais injustement du constat prévu.
- Les informations qu'on veut recueillir sont déjà connues de tous.
- Le programme vient de commencer et n'a pas eu le temps de fonctionner.
- L'autorité veut totalement contrôler la rédaction et la diffusion du rapport.
- On a des raisons de craindre un boycott du recueil de l'information qui invaliderait les résultats.

1.3 Qui sera affecté par, ou impliqué dans l'évaluation?

Il faut bien identifier les personnes et groupes concernés, car ce sont eux qui formuleront les bonnes questions sur lesquelles il faudra ensuite se centrer: Qui demande une évaluation, et pour qui? Qui sera consulté, et sur quoi?

La priorité sera à donner à ceux qui devront ensuite exploiter les résultats: les responsables bien sûr, mais aussi les enseignants et leurs représentants.

Il faut aussi prendre contact avec tous ceux qui manifestent de l'intérêt (auteurs, opposants, associations de parents, etc.) et les questionner ou informer tout au long de l'étude.

1.4 *Quels éléments dans la situation ont des chances d'influencer l'évaluation?*

Certains aspects de la situation qui peuvent être positifs ou négatifs pour l'évaluation ont déjà été évoqués sous 1.B. L'évaluateur doit savoir quelle est l'attitude des différents groupes vis-à-vis de son intervention, pour éviter que son activité soit sabotée, ou au contraire exploitée de façon partielle par certains groupes d'intérêts. Il doit donc connaître l'enjeu de l'évaluation et les sources de difficultés prévisibles.

Il peut arriver aussi que ce soient des éléments favorables ou défavorables de la situation qui expliquent le succès ou l'échec du programme, indépendamment de sa qualité propre. Il peut même arriver que l'évaluation contribue au succès du programme, les participants étant motivés par l'intérêt porté à leur action. L'interprétation ultérieure devra en tenir compte.

Implications pour le plan d'évaluation

- Choix de l'évaluateur (Faut-il un évaluateur interne, ou plutôt externe, ou une équipe formée des deux, pour assurer une évaluation indépendante et crédible?)
- Recueil de l'information (Pourra-t-on faire confiance aux répondants? Seront-ils libres de parler? Pourra-t-on recouper leurs informations par d'autres?)
- Accès à l'information (Des règles administratives ou déontologiques empêchent-elles de consulter certains documents? Obtiendra-t-on les informations critiques pour l'évaluation?)
- Valorisation de l'information (Quelles sources d'information sont discréditées dans ce contexte et donc à laisser de côté? Qui est tenu pour crédible?)
- Rédaction du rapport (Qui va le recevoir d'abord? Qui va l'exploiter? Qui va le diffuser? Qui risque de demander de le modifier? A quelle date devra-t-il être remis?)
- Gestion de l'étude (Avec qui faut-il négocier le contrat? Quelle aide sera donnée dans l'organisation?)
- Évaluation de l'évaluation (Quel rôle veut-on faire jouer à l'évaluateur? Pourra-t-on faire appel à un évaluateur extérieur pour contrôler l'évaluation prévue? Cela est-il souhaitable dans le contexte? Quel évaluateur serait crédible et acceptable?)

1.5 *A quelles questions cruciales doit-on répondre?*

On ne parvient à formuler ces questions cruciales que très progressivement; c'est pourtant d'elles que dépend le plan d'observation et le traitement des résultats. Une démarche en boucles est donc inévitable.

Il faudra traduire des questions trop générales, du type "efficacité du cours" en questions plus précises permettant de leur donner une réponse, comme par ex. "progrès par rapport à l'ancien cours", "objectifs atteints ou non", "degré de satisfaction des participants".

Les différents partenaires ont chacun leurs questions: il faut faire en sorte de leur répondre, mais cela peut accroître beaucoup la charge de travail. Des discussions préalables peuvent contribuer à rapprocher les points de vue sur ce qu'il est plus ou moins important de savoir.

Méthodes pour les faire apparaître

- Analyse conceptuelle des facteurs essentiels de réussite du programme.
- Utilisation d'un modèle théorique de l'apprentissage ou de la formation.
- Recours à l'expérience de programmes analogues.
- Interview de personnes compétentes parmi les groupes concernés
- Analyse de ce qui a déclenché la demande d'évaluation

Méthodes pour les sélectionner

- Prévoir les sources possibles pour obtenir ces informations, les méthodes envisageables, les instruments existants ou à construire, la date à laquelle ces informations devront être disponibles, comment on les analysera, et comment on les interprétera.
- Prévoir les coûts, les difficultés et les retards possibles (éviter les questions trop complexes, en les reformulant de manière à ce qu'elles soient plus accessibles).
- Prévoir aussi qui sera intéressé par le résultat obtenu et dans quelle mesure ce partenaire pourra l'utiliser. Se centrer sur les questions les plus utiles pour fonder des décisions.
- Critères d'évaluation également possibles: Quelle est la question pour laquelle la réponse est la plus incertaine? Qui s'intéresse à cette question? Quelle décision en dépend? Combien coûtera l'investigation? Sait-on déjà quelque chose à ce sujet? L'ensemble des questions est-il équilibré?

1.6 *Le projet d'évaluation a-t-il de bonnes chances de réussir?*

Si oui...

2. Planifier l'évaluation

2.1. *Quels sont les modèles appropriés à la situation?*

Les contacts préalables nécessaires pour répondre aux questions précédentes vont permettre de déterminer quel modèle sera le plus approprié pour apporter une réponse aux questions cruciales.

Dans quel cadre de référence épistémologique se situent ces questions?

- S'il s'agit de porter des jugements sur une réalité extérieure, existant pour elle-même, avec ses relations de cause à conséquence, et indépendamment de l'observateur, on devra appliquer une approche objectiviste, s'inspirant de la démarche expérimentale. Ex.: Des trois méthodes de lecture envisagées, laquelle cause le moins de dyslexies?
- S'il s'agit de fonder des décisions permettant à une institution de satisfaire les groupes sociaux concernés, on se préoccupera moins d'une réalité extérieure insaisissable que de la représentation que s'en font ces groupes et des interprétations qu'ils en donnent. Le cadre de référence sera subjectiviste. Ex.: Les bulletins descriptifs utilisés pour tenir les parents informés des progrès de leurs enfants donnent-ils satisfaction aux divers partenaires dans l'école; parents, enfants, enseignants, et administration scolaire?
- S'il s'agit enfin de permettre à des groupes différents d'arriver à confronter leurs perceptions divergentes pour tendre vers une vision plus complète de la réalité, où chacun tienne compte de la perception des autres, il faudra se situer dans un cadre de référence intersubjectiviste. Ex.: Comment faire

progresser le projet d'établissement de telle école vers une plus grande réussite? Enseignants, parents, et élèves peuvent évaluer le fonctionnement actuel de l'école de façon très différente, mais gagneront à confronter leurs points de vue s'ils doivent s'entendre sur les corrections à y apporter.

Combien d'échelles de valeurs sont prises en compte par le décideur?

On évalue toujours par rapport à son échelle de valeurs propre, mais les décisions dans le domaine social obligent souvent à prendre en compte les valeurs des autres partenaires.

Si les gens voient la réalité différemment, il y a peu de chances que leurs valeurs se rencontrent. Les acteurs sociaux ont naturellement des intérêts divergents du fait que leurs rôles les opposent. Dans un cadre de pensée subjectiviste ou intersubjectiviste, il va de soi que chacun a aussi sa propre échelle de valeurs et que l'évaluateur doit tenir compte de cette multiplicité.

Une échelle de valeurs unique ne peut se concevoir que dans le cadre d'une épistémologie objectiviste. L'économie, par exemple, en offrant de prendre l'argent comme échelle de référence commune, suppose que ce qui se paye est bien réel et qu'un contrat n'est valable que si l'on parle de la même chose.

L'inverse n'est cependant pas nécessaire. Il se peut que des gens s'entendent sur ce que comporte un programme et sur les apprentissages qu'il permet, mais ne valorisent pas ces apprentissages de la même façon. Même dans un cadre de pensée objectiviste, un politicien est obligé de calculer en tenant compte des valeurs de ses adversaires.

On aura donc deux possibilités d'échelles de valeurs (unique, ou multiple) dans l'approche objectiviste, mais une seule (multiple) dans les deux autres.

A quoi l'évaluation doit-elle aboutir?

Vise-t-elle à connaître ou à décider?

Ce à quoi aboutit une évaluation est toujours un rapport, mais cette information peut être relativement générale et abstraite, ou viser un problème particulier plus concret.

Dans le premier cas, le but de l'étude est d'apporter un savoir, utilisable sans doute, mais par n'importe qui, et valable hors contexte, relativement théorique par conséquent. Le décideur qui voudra tenir compte de cette information devra peser le pour et le contre dans son cas particulier. L'évaluation conclura, par exemple, que telle méthode d'apprentissage de l'anglais, si elle est appliquée avant 10 ans, permet d'acquérir un accent correct, mais que cela n'est plus possible au niveau de l'école secondaire. L'évaluateur qui rend compte de ce fait prend en même temps ses distances par rapport au décideur: "aux autorités d'en tirer les conséquences", dira-t-il.

Dans le second cas, le but de l'étude est de fonder une décision qui doit être prise prochainement dans une organisation particulière (que ce soit un décideur unique ou un groupe de négociation). Il s'agit alors de prendre en compte tout le réel, car même un bon programme peut échouer si les circonstances sont trop défavorables, si l'atmosphère sociopolitique de son introduction est trop conflictuelle, par exemple. Il faut alors voir si telle décision est opportune ou non. L'évaluateur essaye dans ce cas d'analyser le problème tel qu'il se pose au(x) décideur(s) et d'élargir ou d'affiner la (ou

leur) connaissance de la réalité. Il essaye de se mettre à la place du décideur pour préparer et soutenir son travail.

Les quatre types envisagés plus haut deviennent donc huit, si l'on distingue à chaque fois les deux cas possibles d'aboutissements visés.

Visé-t-elle à faciliter bilan, gestion ou planification?

L'analyse préalable de la situation doit aussi permettre de dire quelle perspective temporelle est prioritaire: s'intéresse-t-on au passé, au présent, ou à l'avenir, dans un but de certification, de régulation, ou d'orientation?

Il est rare qu'une de ces perspectives soit exclusive. Si l'on s'intéresse au passé, c'est généralement pour en tirer des leçons et pour fonder des décisions préparant l'avenir à plus ou moins long terme. Pourtant, selon la perspective temporelle qui l'emporte, les informations à recueillir et toute la démarche d'observation pourront être différentes. Chaque modèle d'évaluation va donc être "modulé" par cette perspective temporelle.

La méthode est-elle comparative, non comparative, ou multicritères?

Dans le cas particulier où l'évaluateur peut admettre une échelle de valeurs unique, il a alors le choix entre deux méthodes, comparative ou non. Par exemple, un plan d'évaluation de curriculum peut mettre en concurrence un nouveau cours avec plusieurs cours déjà publiés, pour voir lequel est le plus satisfaisant du point de vue des critères préétablis; ou bien le plan d'évaluation peut se contenter de voir si les critères préétablis sont satisfaits par le nouveau cours étudié.

Dans le cas où les échelles de valeurs à prendre en compte sont multiples, la situation est plus complexe. Les méthodes d'évaluation multicritères peuvent aider à combiner des exigences différentes pour fonder la décision.

On arrive ainsi aux dix types de modèles d'évaluation de la figure 13.1, où des noms de modèles déjà publiés et les protagonistes correspondants sont donnés à titre d'exemples indicatifs et approximatifs. On pourrait ajouter de nombreux autres noms pour chaque type.

(Note pour les étudiants: La logique de la typologie reste celle du Ch. 3, mais les exemples de méthodes sont partiellement différents, ou sont classés dans d'autres types: ceci montre que leur catégorisation "après-coup" est incertaine, mais la typologie n'en reste pas moins utile pour décrire les possibilités et les comparer.)

Comme l'indique le bas du tableau, les méthodes de gauche impliquent une préparation relativement rigide, celles de droite une démarche relativement improvisée, en réponse aux problèmes qui apparaissent.

2.2. *Quel doit être le contenu du plan d'évaluation?*

Il correspond nécessairement aux sept étapes du modèle de Michigan, c'est-à-dire qu'en plus des points déjà couverts, il devra traiter du recueil des informations, de leur analyse, du rapport, de la conduite de l'étude et de sa métaévaluation.

Les grilles d'évaluation utilisables à chaque étape devraient donc lui être appliquées de façon "proactive".

2.3. *Comment construire le plan d'observation?*

Par boucles successives, à partir d'un projet mis en discussion, puis en réaction aux éléments nouveaux apparaissant dans l'enquête.

Ce qu'il faut presque toujours décider dès le début:

le but de l'évaluation, les personnes qui seront responsables de l'évaluation, celles qui seront consultées, celles qui la feront, celles qui la recevront et en feront quelque chose, le genre de conclusions auxquelles on veut aboutir ou que l'on veut éviter, les groupes concernés ou comment les identifier, les ressources disponibles, les cadres légaux et administratifs dans lesquels on travaillera, les procédures de contrôle de l'évaluation auxquelles on veut avoir recours et la déontologie à laquelle on se référera, les dates de début et de fin du travail, le calendrier des rapports à faire à chaque audience.

Ce qu'on peut éventuellement laisser pour plus tard:

les types d'instruments et leur contenu, les méthodes de recueil de l'information, les méthodes d'analyse, le contenu du rapport et la façon de le diffuser, les critères d'évaluation.

2.4. Comment s'assurer que le plan est valable?

- Est-il en rapport raisonnable avec l'importance du problème à traiter?
- A-t-il prévu les coûts et anticipé la réaction des audiences à ces coûts?
- A-t-il prévu les problèmes politiques: les réactions des groupes concernés, le contrôle de la publication du rapport, le droit d'intervention des groupes de pression dans la conduite de l'étude en cours de route, les possibilités de renégocier éventuellement le contrat?
- L'acceptation des personnes concernées a-t-elle été vérifiée?
- Les résultats découlant des méthodes prévues semblent-ils probants pour les groupes concernés?

3. Le recueil de l'information

- 3.1. Quelles sortes d'information rassembler?
- 3.2. Quelles procédures, quantitatives et qualitatives, utiliser?
- 3.3. Comment échantillonner?
- 3.4. Comment choisir ou construire ses instruments?
- 3.5. Comment réduire les coûts?

Ces questions techniques ne font pas partie des objectifs du cours.

4. L'analyse et l'interprétation des données recueillies

- 4.1. Comment organiser les données pour pouvoir les traiter?
- 4.2. Comment contrôler les données?
- 4.3. Comment analyser les données?
- 4.4. Comment interpréter les résultats de ces analyses?

Ces questions relèvent des cours de méthodologie et de statistique.

5. Le rapport d'évaluation

Il ne suffit pas de remettre un texte écrit à la fin de la recherche: 1) le compte rendu doit se dérouler aussi oralement, tout au long de l'étude, et se poursuivre même après la remise du rapport. 2) il doit traiter de toute la démarche d'évaluation, depuis ce qui l'a motivée et le contexte dans laquelle elle s'est faite jusqu'aux implications pour l'avenir des résultats obtenus.

5.1. Qui devrait recevoir un rapport?

Il faut faire très attention à l'ordre dans lequel on s'adresse aux divers groupes: les autorités doivent le recevoir avant qu'il soit diffusé à quiconque, mais en même temps, il faut consulter les différents partenaires pour contrôler son

texte et le mettre au point. Il faut s'entendre d'avance sur l'ordre des consultations et sur l'autorité qui décidera de la diffusion.

Un rapport en fin d'étude devrait être fait à tous ceux qui ont participé ou manifesté un intérêt, ou qui devraient en tenir compte à l'avenir. Mais il ne faut pas faire un rapport unique et l'envoyer en masse par la poste. Il faut différencier, voir ce dont chaque audience a besoin et lui donner cette information sous la forme la plus appropriée pour le groupe.

Ne pas oublier de fournir un rapport aussi au personnel qui a collaboré à l'étude, aux groupes interrogés, y compris les élèves ou leurs parents, à ceux qui ont autorisé l'étude, qui l'ont financée, aux collègues qui font un travail analogue, aux bibliothèques universitaires et au Centre de Coordination d'Aarau, éventuellement aux journaux et aux publications scientifiques. Le compte rendu a en effet de nombreuses fonctions pour faciliter l'exploitation des résultats, l'évolution de l'opinion publique, l'acceptation de l'idée d'une démarche expérimentale, etc.

Des rapports devraient paraître dès le lancement, pour l'annoncer, puis des rapports intermédiaires devraient tenir les gens informés de ce qui se passe et des premiers résultats, par exemple 2 à 3 fois par an, puis des rapports internes devraient tenir informées les autorités; les rapports conclusifs devraient paraître sous des formes diverses: rapport principal, rapport résumé, rapport vulgarisé, rapport technique, suite au rapport, conférences de presses, exposés aux associations professionnelles, etc.

5.2. *Comment présenter le rapport?*

Exploiter tous les moyens graphiques et audiovisuels qui existent aujourd'hui pour diffuser de l'information.

"Dire ce qu'on va dire, puis le dire, puis dire ce qu'on a dit".

Éviter dans cet exposé de privilégier un problème, un domaine, un rapporteur, un groupe, etc., mais laisser s'exprimer les minorités.

Privilégier les occasions de discussion, pour redresser les interprétations erronées. Obtenir des feedbacks aussi à propos du rapport écrit avant de le diffuser et même après. Être à disposition pour les séances où le rapport sera discuté dans les groupes concernés (syndicats, etc.).

5.3. *Quand présenter le rapport?*

Avant que soit prise la décision qui a motivé l'étude! La ponctualité est essentielle. Il est possible de diviser le rapport en parties successives, si nécessaire.

6. La gestion de l'étude

6.1. *Qui doit être chargé de l'évaluation?*

Vu le nombre de tâches qu'exige une évaluation, le travail peut être partagé entre plusieurs personnes. La responsabilité principale exige une personne non seulement compétente, mais perçue comme indépendante et acceptable politiquement.

6.2. *Comment établir un contrat?*

Il faut prévoir les points suivants:

- But de l'évaluation.
- Principales questions qu'elle doit aborder.
- Stratégie pour recueillir l'information.
- Procédures pour l'analyser.

- Projet pour le compte rendu: qui recevra un rapport, de quelle nature, quand, quelle possibilité sera offerte d'y réagir.
- Contrôle des biais: ce qui risque d'influencer le recueil des données et ce qu'on prévoit de faire à ce propos.
- Soutien apporté par l'organisation au travail de l'évaluateur.
- Calendrier de l'étude.
- Possibilité de revoir le contrat: comment, quand, dans quelles circonstances.
- Évaluation de l'évaluation: qui viendra superviser la démarche d'évaluation et ses résultats, et à quels moments?
- Budget: sources de financement et bases du payement.

Le contrat peut être établi une fois que le plan d'ensemble a été déterminé.

6.3. Combien devrait coûter une évaluation?

Grossièrement environ 10 % du budget de programme lui-même. Les frais plus élevés que ce pourcentage se justifient parfois par la mise à disposition de résultats supplémentaires utilisables par la suite: formation du personnel, grilles d'évaluation, système de recueil d'informations, etc.

6.4. Comment le travail doit-il être réparti?

Le plan de travail doit prévoir ce qui doit être fait, quand et par qui. Une personne doit avoir la responsabilité d'établir, puis de faire respecter ce plan: c'est généralement la tâche de l'évaluateur.

C'est à lui aussi de prévoir et de gérer les multiples problèmes causés par des changements de situation en cours d'étude: l'important est de veiller à ce que l'étude reste pertinente

7. L'évaluation de l'évaluation

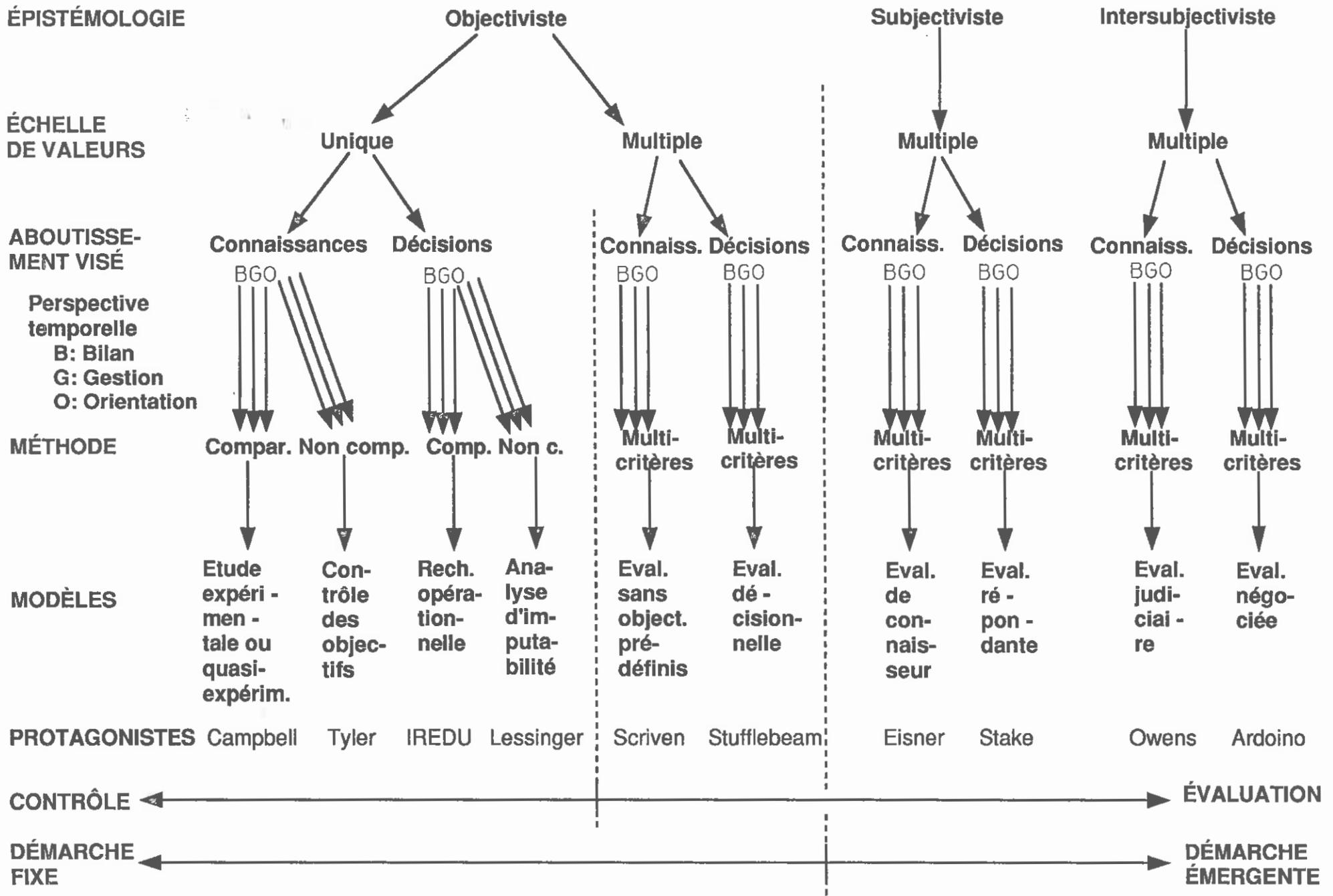
C'est une façon de pratiquer ce que l'on prêche: évaluer ce que l'on fait en cours de route pour pouvoir l'améliorer. C'est une façon aussi d'ajouter de la crédibilité aux résultats de l'étude. Cette réflexion devrait être prévue aux moments critiques: planification de l'étude, interprétation, rapport, mais elle peut être utile à chaque étape.

Elle devrait être faite si possible par quelqu'un qui est extérieur à l'organisation, qui est expert en matière d'évaluation, et qui connaît si possible le domaine du programme. Si personne n'a toutes ces compétences, une équipe interdisciplinaire peut jouer ce rôle.

La démarche à suivre est à nouveau celle des sept étapes ci-dessus.

Les critères permettant d'évaluer si une étude d'évaluation est satisfaisante sont donnés par les "Standards pour l'évaluation de programmes, de projets, et de matériels éducatifs" qui font l'objet du prochain chapitre.

Fig. 13.1 Typologie des modèles d'évaluation



CH. 14 STANDARDS ET DÉONTOLOGIE

1. Le cadre de référence

Des "Standards pour l'évaluation de programmes, de projets et de matériels éducatifs" ont été développés par le "Joint Committee on Standards of Educational Evaluation" en 1981, sous la direction de Stufflebeam. Marc-André Nadeau en a préparé une traduction, reproduite ci-dessous.

Le comité a tenté d'établir un ensemble de standards applicables à toute étude évaluative, que ce soit de programmes, de projets ou de matériels, que ce soit une étude interne ou externe, formelle ou informelle, formative ou sommative, de petite ou de grande dimension. Le comité a de plus déterminé des standards qui encouragent l'utilisation de multiples méthodes et procédures d'évaluation.

Au point de départ, le comité a adopté les définitions suivantes:

- L'objet d'évaluation: c'est ce qui est considéré dans une évaluation, soit un programme, un projet ou un matériel éducatif;

- L'évaluation: investigation systématique de la valeur ou du mérite d'un programme, d'un projet ou d'un matériel;

- Les standards: ils consistent en un ensemble de principes pour mesurer la valeur ou la qualité d'une évaluation, et sur lesquels s'entendent les gens engagés dans la pratique de cette discipline.

Les standards sont considérés comme utilisables par tous ceux qui font de l'évaluation, qu'ils soient ou non des spécialistes.

Les 30 standards établis par ce comité sont regroupés en quatre catégories: d'utilité, de faisabilité, de propriété, et de précision.

2. Les standards d'utilité

Cette catégorie contient des standards propres à guider les pratiques évaluatives afin qu'elles soient informatives et réalisées en temps opportun et qu'elles aient de l'influence. Ces standards exigent que l'évaluateur connaisse les audiences concernées par l'évaluation ainsi que leurs besoins particuliers d'information, qu'il planifie l'évaluation de telle sorte qu'elle puisse répondre à ces informations et qu'il rapporte ces informations de façon claire et en temps opportun.

Globalement, les standards d'utilité sont destinés à assurer une évaluation qui réponde aux besoins d'information pratique des différentes audiences concernées.

2.1. L'identification de l'audience

Selon le Joint Committee, une évaluation implique inévitablement une audience multiple et diversifiée. Les différents groupes de personnes concernées ou affectées par l'évaluation devraient être reconnus afin que leurs besoins puissent être identifiés et connus. Ces audiences incluent les personnes susceptibles d'utiliser les résultats de l'évaluation; ce pourrait également être les individus et les groupes dont le travail est évalué, ou encore les individus susceptibles d'être affectés par les résultats, les organisations communautaires ou le public en général.

2.2. La crédibilité de l'évaluateur

Les personnes responsables de l'évaluation devraient être compétentes et dignes de confiance, afin que leurs conclusions reçoivent un niveau de crédibilité et d'approbation maximal.

Selon le Joint Committee, l'évaluateur sera crédible s'il manifeste les qualités et les caractéristiques jugées nécessaires par le client et les audiences concernées par l'évaluation, par exemple une formation adéquate, une compétence technique, une connaissance réelle, une expérience valable, de l'intégrité et des habiletés particulières en relations publiques. Comme ce sont là des qualités qui se rencontrent rarement toutes chez un seul individu, le comité suggère d'avoir recours à une équipe d'évaluation.

2.3. L'envergure et la sélection de l'information

L'information recueillie devrait être d'une telle étendue et sélectionnée d'une façon telle qu'elle s'adresse aux questions les plus pertinentes, quant à l'objet d'évaluation, et réponde aux besoins et aux intérêts des audiences concernées.

Selon le Joint Committee, l'information est d'étendue suffisante si elle est pertinente et reliée aux objectifs du preneur de décision, si elle est importante pour les audiences concernées et si elle est de portée suffisamment large pour offrir un support valable à un jugement portant sur la valeur ou sur le mérite d'un objet d'évaluation.

L'évaluation ne peut toucher tous les problèmes soulevés par les différentes audiences; l'évaluateur doit sélectionner parmi les problèmes identifiés ceux qui sont d'importance majeure pour les audiences les plus importantes et fournir pour chacune d'elles une information compréhensible et utile. L'objet d'évaluation doit aussi être évalué par rapport à un ensemble de variables importantes telles que: son efficacité, sa faisabilité, ses coûts, sa qualité commé réponse aux valeurs de la société, ses effets secondaires négatifs, etc.

2.4. La valeur dans l'interprétation

Les perspectives envisagées, les procédures et la logique utilisées pour interpréter les résultats devraient être soigneusement décrites afin que les critères sur lesquels reposent les jugements de valeur soient clairs.

Selon le Joint Committee, le concept "valeur" constitue l'essence même du processus d'évaluation; qualifier un objet sur le plan de son utilité, de son importance ou encore de sa valeur générale constitue la principale tâche de la majorité des évaluations. Au coeur de cette activité apparaît la nécessité d'interpréter les informations recueillies en termes de valeurs, que celles-ci soient à caractère quantitatif ou qualitatif, reliées au processus ou aux résultats, de type formatif ou sommatif. Ce processus est cependant complexe et il est susceptible de soulever des controverses. Selon le Joint Committee, l'évaluateur et ses clients doivent mettre beaucoup de sérieux dans le choix d'une approche pour juger les informations en termes de valeurs. De plus, ils doivent expliquer et justifier leur choix.

2.5. La clarté du rapport

Le rapport d'évaluation devrait décrire l'objet évalué et son contexte, ainsi que les objectifs, les procédures et les résultats de l'évaluation, afin que les audiences concernées comprennent rapidement ce qui fut fait, pourquoi cela fut fait, quelles informations furent obtenues, quelles conclusions furent tirées et quelles recommandations furent faites.

Selon le Joint Committee, quel que soit le médium utilisé pour rapporter les résultats de l'évaluation, celui-ci doit être explicite, exempt de détails inutiles et encombrants, et faire usage d'illustrations et de descriptions. Le rapport doit de plus être concis, logique dans son développement, utiliser des termes techniques bien définis, des représentations graphiques et présenter des exemples.

2.6. La diffusion du rapport

Les résultats de l'évaluation devraient être distribués aux clients et aux audiences concernées afin qu'ils puissent les évaluer et les utiliser de façon appropriée. Selon le Joint Committee, l'évaluateur et ses clients doivent s'efforcer de toucher et d'informer toutes les audiences concernées par l'évaluation. Ils doivent ajuster le format et le type de rapport à ces différentes audiences.

2.7. L'à-propos du rapport

Le rapport devrait être produit en temps opportun afin que les audiences puissent utiliser au mieux l'information qui leur est présentée. Selon le Joint Committee, un rapport est jugé à propos et opportun lorsqu'il est remis à un moment où l'audience concernée peut en faire un usage maximum. Un rapport remis en retard perd de sa valeur car son utilisation risque alors d'être fortement diminuée. Cela représente aussi un gaspillage de temps, d'argent et de ressources.

2.8. L'impact de l'évaluation

Les évaluations devraient être planifiées et dirigées de telle sorte qu'elles suscitent et encouragent une continuité chez les membres des audiences concernées.

Selon le Joint Committee, l'impact d'une évaluation représente l'influence que celle-ci peut avoir sur les décisions et les actions des audiences concernées. L'évaluateur ne peut naïvement croire que les améliorations et les changements suggérés vont être immédiatement mis en application une fois le rapport déposé. Il est de sa responsabilité de stimuler et d'encourager les audiences à suivre les recommandations contenues dans le rapport, par exemple apporter des améliorations. A ce titre, il doit jouer le rôle d'agent de changement.

3. Les standards de faisabilité

Cette catégorie contient des standards qui reconnaissent que l'évaluation doit être conduite dans un milieu naturel et qu'elle consomme des ressources. Les standards sont propres à rendre le plan d'évaluation applicable dans un milieu particulier et à exercer un contrôle sur les dépenses en matériel et en personnel.

Les standards de faisabilité sont destinés à assurer une évaluation réaliste, prudente, sobre et diplomate.

3.1. Les procédures pratiques

Les procédures d'évaluation devraient être pratiques afin que les perturbations soient réduites au minimum et que l'information nécessaire puisse être obtenue.

Selon le Joint Committee, les procédures sont les principales actions entreprises pour recueillir et évaluer les informations nécessaires au jugement de la valeur ou du mérite d'un objet d'évaluation. On retrouve, entre autres, les procédures pour la détermination du contrat d'évaluation, pour l'identification des sources de données, pour la détection et l'administration des instruments de mesure, pour le choix des méthodes de collecte, d'enregistrement, d'entreposage et de rappel des données, pour l'identification des procédures d'analyse des données et des moyens de rapporter des résultats.

3.2. La viabilité politique

L'évaluation devrait être planifiée et dirigée avec un effort de prévision des positions différentes que pourraient tenir les groupes ayant des intérêts dans l'évaluation, afin que, d'une part, leur coopération puisse être obtenue et que, d'autre part, les tentatives possibles de ces groupes de réduire l'opération d'évaluation, ou d'influencer ou de mal utiliser ces résultats, puissent être évitées ou contrariées.

Selon le Joint Committee, un groupe d'intérêt est tout groupe qui cherche à influencer une politique dans un sens favorable par rapport à un but ou une préoccupation. Une évaluation est politiquement viable si elle peut poursuivre et atteindre ses buts malgré les pressions exercées et les actions entreprises par différents groupes d'intérêt. L'évaluateur doit prendre des mesures pour assurer la viabilité politique de son travail.

3.3. L'efficacité des coûts

L'évaluation devrait fournir une information ayant une valeur suffisante pour justifier les ressources dépensées.

Selon le Joint Committee, une évaluation est efficace par rapport aux coûts lorsque les avantages qu'on en retire sont égaux ou supérieurs aux coûts encourus. Ces derniers renvoient à la valeur totale, sociale et monétaire, des ressources humaines et matérielles mises à contribution dans l'évaluation; ils incluent autant le coût du matériel et de l'équipement que le temps des participants, des sujets, des volontaires. Les avantages sont représentés par la valeur de tous les résultats provenant de l'évaluation, comme le fait de diagnostiquer correctement une déficience du système et d'y apporter une solution pratique, ou bien de distinguer les services efficaces de ceux qui ne le sont pas, ou encore de réduire le coût de certains services sans en diminuer la qualité.

4. Les standards de propriété

Cette catégorie contient des standards qui reflètent le fait que toute évaluation affecte les être humains qui en sont touchés. Ces standards sont là pour assurer le respect des droits des personnes et des groupes.

Les standards de propriété sont destinés à assurer qu'une évaluation soit dirigée légalement, selon une certaine éthique et avec considération pour le bien-être de ceux qui sont concernés par l'évaluation aussi bien que ceux qui sont touchés par ses résultats.

4.1. L'engagement officiel

Les engagements des parties officielles à une évaluation (ce qu'il y a à faire, comment, par qui, quand) devraient être consignés par écrit afin que ces parties soient obligées d'adhérer à toutes les conditions de l'entente ou puissent entreprendre de les renégocier en bonne et due forme.

Selon le Joint Committee, l'accord écrit constitue une entente mutuelle quant aux attentes et aux responsabilités de l'évaluateur et de ses clients. Il constitue une obligation tant légale qu'éthique de se conformer aux ententes ou encore de les renégocier s'il y a lieu. Le développement d'un contrat permet à l'évaluateur et à ses clients de revoir et de reprendre le plan d'évaluation, de clarifier leurs droits et leurs obligations respectifs.

4.2. Les conflits d'intérêt

Les conflits d'intérêt, souvent inévitables, devraient être discutés ouvertement et honnêtement afin qu'ils ne compromettent pas le processus et les résultats de l'évaluation.

Ils peuvent se manifester autant dans une évaluation interne que dans une évaluation externe; la question n'est pas de savoir comment les éviter, mais plutôt comment se comporter lorsqu'ils se présentent. Selon le Joint Committee,

les conflits d'intérêt peuvent avoir un effet négatif sur l'évaluation par la corruption du processus, des résultats et de leur interprétation. Ils suggèrent plusieurs façons d'attaquer le problème; par exemple, les discussions initiales entre l'évaluateur et les clients peuvent permettre d'identifier et de décrire diverses sources de conflit, de demander les avis de personnes extérieures quant à ces possibilités, d'indiquer par une entente écrite les procédures à respecter pour les éviter, etc.

4.3. La divulgation complète et honnête

Les rapports d'évaluation, oraux et écrits, devraient être ouverts, directs et honnêtes par la divulgation des résultats pertinents, mentionnant les limites de l'évaluation.

Selon le Joint Committee, une évaluation sera défendable à la condition qu'il y ait une divulgation complète et honnête des résultats. Dans ce cas, peut-être que les changements souhaités et suggérés seront entrepris, alors que, dans le cas contraire, l'incertitude risque de miner la crédibilité du rapport.

4.4. Le droit du public à l'information

Les parties officielles à une évaluation devraient respecter et assurer le droit du public à l'information, droit qui, bien sûr, doit être considéré dans les limites des principes et des lois connexes, telles que celles qui touchent à la sécurité du public et au droit à la vie privée.

Le public, ou l'audience concernée, est tout groupe qui moralement et légalement a le droit d'être informé des intentions, des opérations et des résultats d'une évaluation. En l'absence de cette information, les personnes ou les groupes pouvant être affectés par l'évaluation, ne peuvent reconnaître les faiblesses dans les procédures ou dans les données, pas plus qu'ils ne peuvent faire un usage constructif des conclusions.

Le Joint Committee suggère, entre autres, d'inclure dans le projet d'entente ce qui concerne les obligations des parties envers le droit du public à l'information.

4.5. Les droits des personnes

Les évaluations devraient être conçues et dirigées de telle sorte que les droits et le bien-être des personnes soient respectés et protégés.

Selon le Joint Committee, ces droits sont autant ceux qui sont couverts par les lois (consentement de la personne, droit de retrait, droit à la vie privée, etc.) et ceux que l'on retrouve dans les codes de déontologie (respect de l'anonymat, limite quant au temps demandé, protection de la vie et de la santé, etc.), que ceux qui relèvent du sens commun et de la courtoisie (éviter des expériences désagréables, etc.).

4.6. Les interactions humaines

Les évaluateurs devraient respecter la dignité et la valeur humaines dans leurs contacts et leurs échanges avec les personnes associées à une évaluation.

Selon le Joint Committee, l'évaluateur doit faire des efforts particuliers pour respecter les personnes avec qui il entre en interaction dans le cours d'une évaluation, parce que ce sont des être humains qui ont droit au respect, mais aussi pour ne pas mettre l'évaluation et ses résultats en péril.

Selon ce comité, l'évaluateur peut faire preuve de respect, entre autres, en s'intéressant aux préoccupations des participants, en prenant le temps de comprendre les différences culturelles et sociales, en maintenant une communication avec tous.

4.7. L'équilibre du rapport

Le rapport d'évaluation devrait comprendre une présentation complète et claire des forces et faiblesses de l'objet évalué, afin que les forces puissent être utilisées et les problèmes discutés.

Il est important d'être honnête et franc dans la divulgation aussi bien des aspects positifs que négatifs de l'objet évalué.

4.8. La responsabilité financière

Les honoraires versés à l'évaluateur et le coût des ressources impliquées dans l'évaluation devraient refléter des préoccupations d'imputabilité, en plus de la prudence et de la responsabilité déontologique.

Selon le Joint Committee, l'évaluateur pourrait exercer cette responsabilité financière entre autres, en maintenant à jour un dossier des dépenses et des engagements, et un dossier de ses émoluments et du temps consacré à l'évaluation, et aussi en précisant dans le contrat écrit les ententes et les accords sur les sommes allouées aux divers postes, etc.

5. Les standards de précision

Cette catégorie contient les standards qui déterminent si une évaluation a produit de l'information valide et pertinent. Ces standards rappellent que l'évaluateur doit considérer toutes les caractéristiques importantes de l'objet évalué. recueillir de l'information sur chacune d'elles, voir à ce que cette information soit techniquement adéquate et produire des jugements qui soient en relation directe avec les données.

Les standards de précision sont destinés à assurer qu'une évaluation révèle et communique une information techniquement adéquate sur les caractéristiques qui déterminent la valeur ou le mérite de l'objet étudié.

5.1. L'identification de l'objet

L'objet de l'évaluation (programme, projet, matériel) devrait être suffisamment observé afin que la ou les dimensions de l'objet considéré dans l'évaluation puissent être clairement identifiés.

L'objet de l'évaluation, c'est la chose particulière à l'étude, un programme, un projet ou un matériel éducatif.

Ce peut être également une technique ou une façon de faire particulière.

Selon le Joint Committee, il est important pour l'évaluateur d'étudier l'objet en question sur une certaine période de temps pour faciliter le développement et la conduite de l'évaluation, et la rédaction du rapport.

En plus, la description qui en est faite devrait aider les audiences à mieux le connaître dans ses caractéristiques particulières et à lui associer plus facilement les effets positifs et négatifs qu'il engendre.

5.2. L'analyse du contexte

Le contexte dans lequel le programme, le projet ou le matériel est utilisé devrait être examiné d'une façon suffisamment détaillée afin que ses influences probables sur l'objet puissent être cernées.

Le contexte est fait des conditions entourant l'objet d'évaluation, tels sa localisation géographique, son à-propos, le climat politique et social, les activités professionnelles pertinentes en cours, la nature du personnel et les conditions économiques existantes.

Selon le Joint Committee, il est important de connaître ces conditions et autres facteurs pour développer, conduire et rapporter une évaluation réaliste et sensible. Cette connaissance devrait, d'une part, éviter à l'évaluateur de croire à une application plus large que ne le justifient les résultats et, d'autre part, aider les audiences à interpréter l'évaluation, à juger du degré de similitude entre le contexte de l'étude et d'autres contextes.

5.3. La description des buts et des procédures

Les buts et les procédures de l'évaluation devraient être contrôlés et décrits de façon suffisamment détaillée pour être identifiés et évalués. Les buts, ce sont les objectifs de l'évaluation (juger le mérite ou la valeur, contrôler et rapporter le processus d'implantation). Les procédures incluent les diverses façons de collecter, organiser, analyser et rapporter l'information.

Selon le Joint Committee, les buts et les procédures de l'évaluation devraient être précisés et enregistrés à différents moments de l'évaluation: au début, puis à différents moments du déroulement, et enfin à la remise du rapport.

Ces informations sont importantes pour évaluer l'évaluation (métaévaluation), pour aider d'autres évaluateurs à planifier des études évaluatives dans d'autres contextes, pour guider le choix des buts et des procédures à suivre pour une reprise de l'étude, pour aider à la dissémination des techniques utilisées dans l'étude et comme matériel de base dans la formation des évaluateurs.

5.4. Des sources d'information défendables

Les sources d'information devraient être décrites de façon suffisamment détaillée pour qu'on puisse bien juger de la pertinence de l'information. Ces sources incluent les individus, les groupes, les documents audio et vidéo, les

films, etc. et peuvent être utilisées de diverses façons: dans des tests, des observations, des enquêtes et des entrevues.

Selon le Joint Committee, l'évaluateur doit identifier et décrire les sources d'information, les critères et les méthodes utilisés lors de leur sélection, et les moyens pris pour extraire l'information. Il doit de plus décrire les procédures d'échantillonnage utilisées, ainsi que les biais possibles.

5.5. Une mesure valide

Les instruments et les procédures servant à recueillir l'information devraient être choisis ou développés et implantés de telle façon qu'ils assurent la validité de l'interprétation finale.

Le processus de validation consiste à accumuler des données pour justifier l'utilisation et les interprétations résultant de l'usage d'un instrument ou d'une procédure de mesure. La validité d'un instrument est fonction de son utilisation, des questions auxquelles il s'adresse, des conditions de collecte des données, des caractéristiques des personnes à qui il est administré et surtout de l'interprétation des résultats.

Selon le Joint Committee, aucun programme, projet ou matériel ne peut être évalué par une seule et unique mesure; il est donc important, pour bien évaluer un programme, de faire appel à toutes les variables importantes.

De plus, comme les instruments de mesure et les utilisateurs ne sont pas parfaits, l'usage de mesures multiples peut pallier les déficiences des uns et des autres. Selon le Joint Committee, les évaluateurs devraient voir à mesurer de multiples résultats et faire usage de nombreux instruments de mesure (chacun de ceux-ci devant être valide), y compris les questionnaires, les entrevues, les échelles de mesure, les observations, les tests et les descriptions.

5.6. Une mesure fidèle

Les instruments et les procédures servant à recueillir l'information devraient être choisis ou développés, et utilisés, de telle façon que l'information obtenue possède un degré de fidélité suffisant pour l'usage attendu.

Une mesure fidèle est celle qui fournit des informations constantes quant à la caractéristique étudiée.

Selon le Joint Committee, les évaluateurs devraient choisir un ou des instruments de mesure qui ont un niveau de fidélité acceptable pour leur utilisation particulière et pour la ou les situations particulières dans lesquelles ils sont utilisés. Ils devraient également vérifier et rapporter le degré de fidélité des mesures prises dans leur étude évaluative.

5.7. Le contrôle systématique des données

Les données recueillies, traitées et rapportées dans l'évaluation, devraient être révisées et corrigées afin que les résultats ne soient pas entachés d'erreurs.

Le contrôle systématique des données signifie que toutes les mesures possibles seront prises pour que les données utilisées soient exemptes d'erreurs humaines.

Le Joint Committee suggère que les évaluateurs mettent en place des mesures de contrôle et de vérification afin d'éviter qu'il ne se glisse des erreurs humaines (autant des répondants que de l'évaluateur ou des codificateurs et analystes) lors de la collecte, de l'enregistrement, de la collection, du traitement et de l'analyse des données.

5.8. L'analyse de l'information quantitative

Afin d'assurer des interprétations défendables, l'information quantitative devrait être convenablement et systématiquement analysée.

L'information quantitative est composée de faits et de prétentions (âge, caractéristiques socioéconomiques, performance, attitudes, comportements, caractéristiques de l'objet d'évaluation, etc.) représentés par des chiffres et, pour des fins d'analyse, compilés, organisés, manipulés et validés afin de répondre à certaines questions relatives à un objet donné.

Selon le Joint Committee, l'analyse quantitative peut aboutir à différentes conclusions et il appartient aux évaluateurs de faire en sorte que les méthodes d'analyse utilisées ne conduisent pas à des interprétations abusives et (ou) fautives et (ou) fallacieuses. Les évaluateurs doivent être en mesure de justifier et de défendre leur approche méthodologique, les postulats sur lesquels elle s'appuie, leurs calculs et leurs conclusions.

5.9. L'analyse de l'information qualitative

Afin d'assurer des interprétations défendables, l'information qualitative devrait être convenablement et systématiquement analysée.

On entend par information qualitative les faits et les interprétations qui épousent la forme narrative plutôt que numérique. Il s'agit de trouver des réponses à des questions concernant l'objet d'évaluation. Contrairement à l'analyse quantitative, on peut ici donner de la perspective et de la profondeur à des données.

Selon le Joint Committee, l'information qualitative peut provenir de différentes sources (des entrevues, structurées ou pas, des observations, participantes ou pas, des documents et des dossiers, etc.); elle peut être intentionnelle ou non; elle peut se rapporter aux décisions, aux objectifs, aux plans et procédures, aux résultats, etc.; elle peut être notée sous la forme de descriptions, d'argumentations logiques, d'interprétations ou encore d'impressions subjectives.

Les évaluateurs doivent faire des efforts sérieux pour éviter des conclusions abusives et (ou) fautives et (ou) fallacieuses, et ce, en choisissant des méthodes d'analyse appropriées, en procédant à des contre-vérifications et en s'assurant de la validité et de la fidélité des instruments et des procédures de mesure.

5.10. Des conclusions justifiées

Les conclusions d'une évaluation devraient être explicitement justifiées afin que l'audience puisse les évaluer.

Elles se présentent comme des jugements et des recommandations, et doivent être défendues et défendables, d'une part, parce qu'elle peuvent être fautives et, par conséquent, mener à des actions inappropriées et, d'autre part, parce qu'elles peuvent être ignorées en l'absence d'information. Selon le Joint Committee, pour être défendables, les conclusions doivent s'appuyer sur une grande logique et sur une information appropriée. Elles seront défendues si elles sont accompagnées d'une description des procédures, de l'information et des postulats de base, de même que d'une discussion des interprétations optionnelles et des raisons de leur rejet, s'il y a lieu.

5.11. L'objectivité du rapport

Les procédures d'évaluation devraient assurer une garantie contre la distorsion des résultats et des rapports d'évaluation, distorsion qui peut être causée par les sentiments et les préjugés personnels des parties en cause.

Selon le Joint Committee, les rapports d'évaluation peuvent être biaisés de plusieurs façons: ils peuvent ne pas représenter les diverses perspectives ou ne pas en faire état, ils peuvent être conçus de façon à tromper, à déformer ou cacher des faits, à entériner des décisions déjà prises, etc. Certaines distorsions peuvent être le résultat de l'ignorance ou de l'imprudence de l'évaluateur. D'autres peuvent être le résultat de pressions exercées par les clients, les audiences concernées ou les bailleurs de fonds. L'évaluateur doit veiller à l'intégrité et à l'objectivité du rapport. Ce sera possible dans la mesure où le rapport donnera une présentation impartiale des faits, et évitera de prendre position en faveur d'un groupe au détriment d'un autre.

Marc-André Nadeau, *L'évaluation de programme, Théorie et pratique*,
Québec: Les Presses de l'Université Laval, 1988, p.141-155.

14. B. QUESTIONS AUX GROUPES

Vous recevez deux documents: d'une part une présentation sommaire de la procédure d'évaluation de l'Université de Fribourg, avec ses objectifs généraux et la liste des questions posées, d'autre part une traduction des Standards acceptés par les évaluateurs américains en matière d'éducation.

Votre tâche est de mettre en rapport ces deux documents pour présenter des recommandations au Comité d'Évaluation (notamment au délégué étudiant) concernant la procédure envisagée:

- dangers prévisibles à éviter;
- façons de faire à privilégier;
- qualités à rechercher, etc.

Sentez-vous très libres de faire des commentaires en ce stade tout à fait préliminaire de l'étude: le but de l'exercice est d'abord éducatif. Il s'agit pour vous de découvrir la pertinence des Standards par rapport à une situation concrète: ces règles, toutes générales qu'elles soient, peuvent vous aider à imaginer et planifier une étude d'évaluation.

Vu l'ampleur du travail, ne traitez qu'une partie des Standards:

- soit les standards d'utilité et de faisabilité,
- soit les standards de propriété,
- soit les standards de précision,

et relevez les commentaires qu'ils vous suggèrent pour le cas de Fribourg.

Trente minutes avant la fin du cours, chaque groupe donnera la parole à un rapporteur (à désigner dès le début) pour qu'il explique les réflexions et propositions de son groupe.

Veillez prendre note de ces réflexions qui pourraient être utiles à votre délégué.

RÉPONSES DES GROUPES

2. Les Standards d'utilité

2.1. L'identification de l'audience

L'évaluation doit répondre aux besoins des instances concernées : internes (enseignants, universitaires, étudiants, etc.) et externes (maîtres d'école primaire, secondaire, et de gymnase ainsi que des organisations diverses).

L'audience est multiple et diversifiée. Elle ne comprend pas seulement des étudiants et des enseignants, mais aussi des politiciens, des associations professionnelles, l'Église catholique qui subventionne l'Université.

2.2. La crédibilité de l'évaluateur

L'évaluateur doit être compétent, bien formé. Il doit avoir une bonne expérience, des compétences techniques et de l'habileté dans les relations publiques.

Le groupe d'évaluation doit comporter le plus grand nombre de participants possibles (par exemple des pédagogues).

2.3. L'envergure et la sélection de l'information

Il faut sélectionner les questions en fonction de l'efficacité et de la qualité des services, des effets secondaires négatifs susceptibles de se produire. Il faut se centrer sur ce qui ne va pas bien.

Jean Cardinet ajoute qu'il serait utile de juger les questions de la liste prévue pour l'évaluation en fonction des critères mentionnés dans cette rubrique des standards.

2.4. La valeur dans l'interprétation

La notion de valeur est ambiguë, compliquée et sujette à controverse. Qu'est-ce qui est important ? La recherche ou l'enseignement ? Il est nécessaire de réfléchir au référentiel.

Pour Jean Cardinet le Joint Committee rejoint l'exigence de référentiel de Gérard Figari. Par rapport à quoi juge-t-on, évalue-t-on ? C'est là-dessus qu'il faut mettre le doigt.

2.5. La clarté du rapport

Il faut utiliser toutes les présentations possibles : tableaux, graphiques, moyens audiovisuels, etc., à l'intention des publics cibles particuliers.

2.6. La diffusion du rapport

Plusieurs rapports différents doivent être rédigés afin d'être adaptés à chaque public cible.

4. Les standards de propriétés

4.1. L'engagement officiel

L'engagement doit être négociable et consigné par écrit. La partie évaluée n'est pas simplement passive. Un contrat doit être établi entre le Conseil supérieur et les personnes qui vont réaliser l'évaluation; l'Université qui va subir l'évaluation devrait faire valoir ses droits.

4.2. Les conflits d'intérêt

Il faudrait prévoir la possibilité qu'il y ait des conflits et prévoir leur résolution. Il faut une discussion initiale pour identifier les conflits possibles, non pas pour les éviter, mais au contraire pour les gérer.

Dans l'évaluation de l'Université de Fribourg, un contrat est par exemple nécessaire, car l'éventualité de conflits avec le Rectorat n'est pas à exclure.

4.3. La divulgation complète et honnête

Les rapports de l'évaluation doivent être écrits et oraux ! Tous les résultats positifs ou négatifs doivent être présentés.

4.4. Le droit du public à l'information

Les résultats doivent pouvoir être consultés par tout le monde. Ce point, par exemple, n'est pas explicité dans le document. Il faut toutefois prendre des précautions pour ne pas créer plus de problèmes que l'on en résout.

4.5. Les droits des personnes

Les gens ont droit de connaître ce qui est dit; en retour, on doit respecter leurs avis; ils ont droit à la parole, par exemple, ils ont le droit de réagir pour se défendre.

Il faut cependant veiller au problème de l'anonymat. Si une seule personne est responsable d'un Institut, toute chose n'est pas bonne à dire, car tout le monde peut aisément repérer la personne.

Bien sûr, l'évaluation n'est pas faite pour dire uniquement du bien. Mais il faut prêter attention à la forme du rapport; ne pas être injurieux, ne pas chercher le scandale mondain.

4.6. Les interactions humaines

Le point de vue des évaluateurs doit être respecté. Mais on veillera à présenter plusieurs points de vue.

4.8. La responsabilité financière

Les coûts sont difficiles à prévoir. Il faut cependant prêter attention au rapport coûts/somme mise à disposition.

Un moyen pour fixer les coûts de l'évaluation consiste à déterminer trois parts : la première estimation concernerait les trois premiers mois, on procéderait à une deuxième estimation ensuite, puis à la fin de la deuxième phase, il serait possible d'estimer les moyens financiers qui restent.

5. Les standards de précision

5.1. L'identification de l'objet

Quel est l'objet évalué dans le cadre de l'évaluation de l'Université de Fribourg est bien sûr une des premières questions à se poser.

5.2 L'analyse du contexte

Dans ce qu'on a lu, on ne mentionne pas de manière détaillée le contexte. Par exemple, rappeler "l'état actuel des exigences scientifiques, de la culture, et des conditions socio-économiques" ne constitue pas vraiment une connaissance détaillée du contexte en question.

Il serait bon de le rappeler aux délégués des étudiants, par exemple.

5.3. La description des buts et des procédures

Pour pouvoir évaluer l'évaluation elle-même, il serait bon de tenir un journal, de dire comment on arrive à tels résultats, surtout si ceux-ci n'avaient pas été prévus; de donner des indications sur les instances qui ont constitué le comité d'évaluation, etc.

5.4. Des sources d'informations défendables

Il faut varier les sources d'informations et les moyens pour récolter les données afin de les rendre pertinentes (procédé de triangulation). Par exemple, dans le cas de l'Université de Fribourg, interroger, demander le point de vue des étudiants, peut servir à contrôler les informations recueillies auprès des professeurs, et inversement.

5.5. Une mesure valide

Les interprétations sont validées si on y aboutit par le moyen de mesures multiples.

5.6. Une mesure fidèle

Il faudrait faire attention, par exemple, à ne pas prendre les informations uniquement sur la dernière année, mais il faudrait choisir au contraire une période assez large.

5.7. Le contrôle systématique des données

Pour le relevé des résultats chiffrés, demander un contrôle, dans certains cas une contre-expertise, peut éviter des erreurs matérielles.

5.8 et 5.9 L'analyse de l'information quantitative et l'analyse de l'information qualitative

Ces points concernent les problèmes de l'interprétation. Il faut éviter les interprétations abusives ou erronées.

5.10. Des conclusions justifiées

Les conclusions avancées doivent être mises par écrit. Leur comparaison avec celles obtenues par d'autres juges, d'autres régions, devient alors possible.

5.11. L'objectivité du rapport

Il faut éviter de prendre position pour l'un ou l'autre groupe concerné par l'évaluation.

EVALUATION DE L'UNIVERSITE DE FRIBOURG

MISE EN PLACE ET DEROULEMENT.

DEROULEMENT DE LA PROCEDURE (SOMMAIRE)

1. Le Rectorat arrête les principaux points à évaluer en priorité, sous la forme d'une *liste de questions*. L'évaluation consistera à obtenir des réponses à ces questions. (cf. document eval-05).
2. Un arrêté du Conseil d'Etat *institue le Conseil Supérieur de l'Evaluation*, et le charge de définir les objectifs de l'évaluation, sur la base des propositions élaborées par le Rectorat.
3. Déblocage des *moyens financiers*: (1) montants mis à disposition par le CDU; (2) montant prévu dans le cadre de l'enveloppe budgétaire de l'Université.
4. Le Conseil supérieur de l'évaluation *désigne le président du Comité d'évaluation, ainsi que le bureau*. En même temps, il institue le Comité d'évaluation ("ComEv") et *définit son mandat* (formulation définitive des objectifs).
5. Mise en place des *commissions facultaires* par le ComEv, et début des investigations.
6. *Déroulement* de l'évaluation, en 3 phases: (a) le ComEv recueille les données utiles et élabore une première analyse; (b) les commissions facultaires prennent position sur cette analyse (critique, ré-interprétation des résultats, compléments d'analyse...); (c) le ComEv rédige un rapport final qui intègre ces observations.

Rem. La procédure est prévue en forme de débat contradictoire, afin de concilier deux impératifs: (1) donner au ComEv la maîtrise des opérations, donc une grande liberté d'action, permettant d'imposer méthodes, critères, rythmes de travail; (2) garantir que les intéressés pourront faire valoir leurs points de vue. Pour cela, on prévoit une "navette" entre ComEv et commissions facultaires. Durant la phase (a), le ComEv a la direction des opérations; il utilise les groupes de base comme sources de renseignements, et élabore en toute indépendance ses propres conclusions. La phase (b) permet aux Facultés de prendre position sur ces conclusions. La phase (c) rend au ComEv l'autorité de conclure.

7. Remise du rapport au Conseil Supérieur de l'Evaluation. Celui-ci en diffuse le contenu auprès des instances intéressées, et met tout en oeuvre pour aboutir à la *réalisation des réformes* proposées.

EVALUATION DE L'UNIVERSITE DE FRIBOURG:

OBJECTIFS

0. PRELIMINAIRES.

Ce document définit les **buts** du programme d'évaluation de l'Université de Fribourg (1993-95).

Ces buts ont été formulés à la lumière d'une **consultation préalable**, organisée auprès des Facultés et Sections, qui a permis d'établir un inventaire des principaux problèmes auxquels se trouve actuellement confronté le corps enseignant.

La présente liste constitue un **état idéal, maximal**, des objectifs à atteindre. Elle devra être adaptée en fonction des moyens disponibles, des contraintes temporelles, et de la nature des unités évaluées.

1. BUTS GENERAUX.

L'évaluation a pour but d'apporter une réponse à **trois questions** principales:

1. Comment l'université peut-elle remplir plus efficacement les missions d'enseignement et de recherche qui lui sont fixées, compte tenu de l'état actuel des exigences scientifiques, de la culture, et des conditions socio-économiques?
2. Les moyens, et leur répartition, sont-ils appropriés aux tâches à accomplir? (Quels sont les secteurs sous-dotés? Sur-dotés?)
3. Quelles améliorations faut-il apporter aux structures et aux procédures de gestion pour les rendre plus efficaces?

2. INVENTAIRE DETAILLE.

Ces trois objectifs fondamentaux peuvent être divisés en divers sous-buts. D'où une liste de questions détaillées, auxquelles il conviendra de répondre pour chaque unité d'enseignement et de recherche évaluée:

1. Orientations generales de l'enseignement et de la recherche.

1.1. Enseignement.

1.1.1. Les contenus d'enseignement correspondent-ils à l'état actuel des connaissances dans la branche? Existe-t-il des lacunes ou des sous-domaines non enseignés, et pourquoi?

1.1.2. L'enseignement de certaines matières pourrait-il être abandonné, ou réduit à l'état d'options?

1.1.3. Y a-t-il des traditions historiques qui déterminent l'orientation actuelle de l'enseignement? Sont-elles un avantage ou un inconvénient?

1.1.4. Quelle est la proportion entre (a) formation à la spécialité (branche principale), (b) formation complémentaire au service d'autres disciplines (branche secondaire), (c) conservation de savoirs "confidentiels", (d) autres fonctions?

1.1.5. Des enseignements interdisciplinaires sont-ils donnés dans le cadre de la branche? De quelle nature sont-ils? De quel ampan?

1.1.6. Les contenus d'enseignement sont-ils appropriés aux besoins des étudiants et à ce que l'on connaît du marché de l'emploi? Garantissent-ils des facultés d'adaptation ou des reconversions ultérieures?

1.1.7. La proportion entre formation scientifique fondamentale d'une part, formation appliquée, professionnelle ou pratique d'autre part est-elle judicieuse?

1.1.8. Les choix de branches faits par les étudiants sont-ils cohérents? Expriment-ils certains stéréotypes? Faut-il faire évoluer ces profils?

1.1.9. Le système d'examens est-il approprié?

1.1.10. Existe-t-il une articulation rationnelle entre les 1er, 2ème et 3ème cycles? Les plans d'études peuvent-ils être mieux structurés?

1.1.11. Une articulation entre formation initiale et formation continue est-elle prévue? Comment l'organiser?

1.2. Recherche.

1.2.1. Quelles sont les activités de recherche menées dans la branche? Quelles formes prennent-elles? Quelle est la proportion entre tâches d'enseignement et de recherche?

1.2.2. Les résultats de la recherche sont-ils publiés? Quelle est leur audience? Le personnel enseignant a-t-il le temps d'écrire?

1.2.3. Quelles sont les retombées de la recherche sur l'enseignement?

1.2.4. Quelles sont les répercussions de l'enseignement sur la recherche?

1.2.5. Quels sont les autres impacts de la recherche?

1.2.6. Y a-t-il des spécificités propres à l'Université de Fribourg, en ce qui concerne la recherche dans la branche? Quel est son degré d'originalité?

1.3. Services.

1.3.1. Quelles activités de service l'Université assure-t-elle? Quels autres services pourraient être rendus?

2. Moyens.

2.1. Taux de recouvrement: le nombre d'enseignants (professeurs & cadres intermédiaires) est-il suffisant par rapport au domaine de connaissances à couvrir?

2.2. Qualification des personnels: correspond-elle aux contenus à enseigner? Emploie-t-on du personnel sous-qualifié, sur-qualifié, qualifié pour autre chose?

2.3. Taux d'encadrement: le nombre d'enseignants (professeurs & cadres intermédiaires) est-il suffisant au regard du nombre des étudiants?

2.4. Qualité de l'encadrement: les méthodes d'enseignement sont-elles appropriées (trop ou trop peu scolaires, interactives ...)?

2.5. Bilinguisme: quels sont ses avantages et ses coûts? Existe-t-il des disparités entre français et allemand dans la branche? A quelles conditions peut-on y délivrer des diplômes bilingues? Le statut de l'italien pose-t-il problème?

2.6. Moyens administratifs: sont-ils suffisants? Le temps consacré par les professeurs à la gestion et à l'acquisition de moyens est-il équilibré?

2.7. Comment réunir les données quantitatives permettant de contrôler les flux d'étudiants ? (durée des études, drop-out, nombre de licences et autres diplômes, nombre d'heures d'enseignement par branche...)

3. Structures.

3.1. L'organigramme actuel des Facultés, Sections, Instituts... est-il adapté aux besoins de l'enseignement et de la recherche? Impose-t-il ou favorise-t-il des cloisonnements nuisibles entre disciplines? Quels sont les dégroupements et regroupements d'unités à envisager?

3.2. La coopération scientifique avec les branches les plus voisines peut-elle être améliorée? Comment?

3.3. Comment favoriser la coopération interfacultaire et interuniversitaire dans la branche?

3.4. Moyens administratifs: sont-ils suffisants? Le personnel administratif des décanats, sections ou autres est-il suffisamment qualifié au regard des tâches à accomplir? Les rapports hiérarchiques et de service sont-ils efficaces?

3.5. L'organigramme actuel de l'Université est-il adapté aux besoins de la gestion et de l'administration? Comment les procédures administratives peuvent elles être simplifiées? Accélérées? Comment éviter aux enseignants les surcharges de tâches administratives?

3.6. Les services généraux (bibliothèques, informatique, immatriculation, etc.) répondent-ils efficacement aux besoins? Quels services généraux n'existent pas et seraient à créer?

CH. 15B. POUR ÉVALUER LES ÉVALUATIONS

1. PRÉPARATION

1.1. La question est-elle explicite?

- 1.1.1. *A-t-on étudié la situation de départ (du point de vue pédagogique et institutionnel) pour savoir dans quelle optique analyser les problèmes?*
- 1.1.2. *L'étude a-t-elle été planifiée en vue d'une décision à prendre?*
- 1.1.3. *A-t-on explicité les questions, les attentes, et les référentiels d'évaluation des divers partenaires?*
- 1.1.4. *La formulation de la question de recherche a-t-elle été négociée?*
- 1.1.5. *A-t-on obtenu l'accord des groupes politiques sur le principe de cette recherche?*

1.2. La question est-elle judicieuse?

- 1.2.1. *L'hypothèse méritait-elle une telle recherche (hypothèse trop évidente, ou trop peu fondée théoriquement)?*
- 1.2.2. *Le plan de la recherche a-t-il été réfléchi et mis au point à l'avance?*
- 1.2.3. *La faisabilité et le coût ont-ils été estimés?*
- 1.2.4. *Les inconvénients ont-ils été minimisés, pour les élèves et les enseignants, dans la prise d'information envisagée?*
- 1.2.5. *La réponse pouvait-elle être prise en compte par les partenaires (politiquement et économiquement)?*
- 1.2.6. *La question en cachait-elle une autre?*
- 1.2.7. *La réponse pouvait-elle aider à prendre la décision envisagée?*

1.3. Un contrat a-t-il été établi avec le mandant au départ?

- 1.3.1. *sur la gestion administrative de la recherche?*
- 1.3.2. *sur le coût de l'étude et sa gestion financière?*
- 1.3.3. *sur le délai de remise du rapport?*
- 1.3.4. *sur le contenu du rapport?*
- 1.3.5. *sur la publication du rapport?*

1.4. Le projet est-il éthiquement acceptable?

- 1.4.1. *concernant le respect des droits des individus*
- 1.4.2. *concernant le respect des droits du public.*

2. RÉALISATION

2.1. Le modèle est-il approprié à la question primordiale?

- 2.1.1. *Cherchait-on une réponse universelle, ou locale?*
- 2.1.2. *Acceptait-on une ou plusieurs échelles de valeurs?*
- 2.1.3. *L'orientation de la question était-elle théorique ou pratique?*
- 2.1.4. *L'évaluation était-elle comparative, ou non?*

2.2. Les informations recueillies sont-elles suffisantes?

- 2.2.1. *Quantitativement:*
 - 2.2.1.1. *La taille des échantillons permettait-elle de conclure?*
 - 2.2.1.2. *Le mode d'échantillonnage permettait-il de généraliser?*

- 2.2.1.3. **La source des informations a-t-elle été précisée suffisamment pour permettre d'en reprendre d'autres?**
- 2.2.1.4. **Le traitement des données a-t-il été contrôlé?**
- 2.2.2. *Qualitativement:*
 - 2.2.2.1. **A-t-on contrôlé la fidélité des dispositifs de mesure utilisés?**
 - 2.2.2.2. **A-t-on contrôlé les biais possibles (positifs et négatifs) pouvant affecter les résultats?**
 - 2.2.2.3. **A-t-on tenu compte de l'influence du contexte?**
- 2.3. **La conclusion est-elle valide?**
 - 2.3.1. *A-t-on mesuré les performances avant et après traitement?*
 - 2.3.2. *A-t-on comparé groupe expérimental et groupe contrôle?*
 - 2.3.3. *A-t-on analysé correctement l'ensemble des résultats?*
 - 2.3.4. *A-t-on étudié la façon dont le traitement a été appliqué?*
 - 2.3.5. *En a-t-on tenu compte pour corriger l'application?*
 - 2.3.6. *A-t-on examiné les effets directs et indirects du curriculum?*
 - 2.3.6.1. **sur les élèves?**
 - 2.3.6.2. **sur les enseignants?**
 - 2.3.6.3. **sur les administrateurs?**
 - 2.3.6.4. **sur les parents?**
 - 2.3.7. *A-t-on comparé les effets pour différentes catégories d'élèves (niveaux socio-économiques, sexes, nationalités...)*
 - 2.3.8. *A-t-on écouté les réactions des partenaires?*
 - 2.3.9. *A-t-on relevé tous les événements relatifs à l'expérience?*
 - 2.3.10. *A-t-on cherché à interpréter l'histoire de la recherche?*
 - 2.3.11. *A-t-on explicité la démarche d'interprétation suivie, en même temps que son résultat?*
- 2.4. **La conclusion est-elle généralisable?**
 - 2.4.1. *Les conditions étaient-elles particulières?*
 - 2.4.2. *La recherche a-t-elle été influencée par l'attitude des participants?*
 - 2.4.3. *La recherche a-t-elle des implications théoriques générales?*
 - 2.4.4. *La recherche a-t-elle des implications méthodologiques?*
 - 2.4.5. *La recherche a-t-elle des implications pratiques?*

3. RÉSULTATS

- 3.1. **Le rapport est-il compréhensible par les intéressés?**
 - 3.1.1. *A-t-on diversifié les formes de rapport?*
 - 3.1.2. *S'est-on adressé directement aux diverses audiences, en cours de recherche, à la fin, et ultérieurement?*
- 3.2. **La réponse est-elle acceptable politiquement?**
 - 3.2.1. *Risque-t-elle d'être en désaccord avec les finalités de l'école?*
 - 3.2.2. *Risque-t-elle d'être en désaccord avec la politique des autorités du moment?*
 - 3.2.3. *Renforce-t-elle uniquement le pouvoir de ses mandants?*
 - 3.2.4. *Menace-t-elle les intérêts de certains groupes? Ces derniers ont-ils été consultés?*

3.3. La réponse est-elle utilisable?

3.3.1. *A-t-on répondu au moment opportun?*

3.3.2. *L'étude a-t-elle eu des implications pratiques:*

3.3.2.1. directes?

- 3.3.2.2. indirectes?

3.3.3. *L'étude a-t-elle eu des effets indirects:*

3.3.3.1. positifs?

3.3.3.2. négatifs?

3.3.4. *L'étude peut-elle être mise en rapport avec d'autres?*

3.3.5. *Les décisions ultérieures ont-elles tenu compte de l'étude?*

3.4. Le coût est-il raisonnable?

3.4.1. *coût de la recherche elle-même?*

3.4.2. *coûts et gains indirects de la recherche?*

3.4.3. *coût ultérieur de l'application du curriculum?*

3.4.4. *comparaison avec le coût attendu d'autres curriculums concurrents?*