

L'élargissement de l'évaluation

Autor(en): **Cardinet, Jean**

Objektyp: **Article**

Zeitschrift: **Bildungsforschung und Bildungspraxis : schweizerische Zeitschrift für Erziehungswissenschaft = Éducation et recherche : revue suisse des sciences de l'éducation = Educazione e ricerca : rivista svizzera di scienze dell'educazione**

Band (Jahr): **1 (1979)**

Heft 1

PDF erstellt am: **15.09.2024**

Persistenter Link: <https://doi.org/10.5169/seals-786448>

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

L'élargissement de l'évaluation

Jean Cardinet

La recherche pédagogique devrait pouvoir contribuer à l'adaptation continue du système scolaire; ceci implique qu'elle réponde en priorité aux questions des responsables, en leur apportant l'information qu'ils demandent pour fonder leurs décisions. Il en résulte un déplacement d'accent, par rapport à la recherche scientifique habituelle. Ce n'est plus le chercheur qui évalue, mais les personnes engagées dans l'expérience. L'approche nouvelle de l'évaluation repose finalement sur une conception de la science non plus positiviste, mais fondée sur la compréhension des phénomènes, envisagés dans leur cadre de référence subjectif. L'expérimentation se transforme en recherche-action où le chercheur est lui-même un des éléments de la situation qu'il observe. L'objectivité devient accord intersubjectif des partenaires dans la réforme.

Les conceptions pédagogiques ne sont plus aujourd'hui ce qu'elles étaient il y a dix ans. Chacun s'en rend compte. Ce que l'on sait peut-être moins, c'est que les principes de la recherche pédagogique sont également en pleine transformation. Deux manifestations internationales récentes viennent de le souligner. Il s'agissait d'abord du Symposium de Kessel sur «L'Évaluation des projets pilotes de réforme scolaire», organisé par le Conseil de l'Europe. Ce fut ensuite la Conférence de Liège sur les «Nouvelles méthodes d'évaluation des programmes d'enseignement», préparée par le CERI-OCDE. Bien que les idées présentées dans ces deux contextes se distinguent sur bien des points, elles se rencontrent sur leur thèse fondamentale: la méthode expérimentale, utilisée dans les laboratoires, ne suffit pas aux besoins de la pédagogie.

La crise est apparue d'abord aux États-Unis. Une loi a été votée, exigeant que les programmes éducatifs nouveaux, financés par l'administration, fassent l'objet d'une évaluation scientifique. Les recherches et les publications n'ont pas tardé à fleurir. Leurs fruits, pourtant, ont été décevants. Les responsables des programmes nouveaux n'ont pas reçu l'appui qu'ils escomptaient pour améliorer leur action. Les conclusions venaient trop tard. Elles étaient trop partielles. Sous prétexte d'objectivité, elles négligeaient souvent le sentiment subjectif de ceux qui se trouvaient engagés dans ces réformes. Elles mettaient ainsi en cause l'utilité même des réformes, alors que c'était peut-être le manque de finesse des instruments utilisés pour les évaluer qu'il aurait fallu incriminer.

Les discussions passionnées soulevées par les études d'évaluation ont eu au moins un résultat positif: chercheurs et administrateurs ont commencé à dialoguer; la différence de leurs perspectives est apparue au grand jour et, avec elle, la nécessité de repenser aussi l'évaluation.

Certaines solutions commencent à apparaître. Les deux manifestations citées ci-dessus avaient pour but de les faire connaître. On avait invité à la fois des universitaires et des responsables scolaires pour qu'ils poursuivent le dialogue commencé Outre-Atlantique. L'objectif était de déterminer une méthodologie de l'évaluation des expériences pédagogiques qui satisfasse les exigences des uns et des autres.

Les conceptions proposées vont être décrites ci-dessous. La présentation suivra un ordre à la fois logique et chronologique. Il faut s'entendre d'abord sur le but de l'évaluation. Cela aura des conséquences sur le choix des participants à l'évaluation, sur l'objet à évaluer et sur la nature des instruments destinés à recueillir les données. On aura alors formulé une nouvelle méthodologie fondée sur de nouveaux critères. Il faudra aussi repenser la transmission et l'utilisation de l'information nouvelle qu'on aura élaborée.

Pour chacune des questions précédentes, on opposera une conception étroite de l'évaluation et une conception élargie. Cette comparaison sera naturellement plus caricaturale que réaliste. Elle aura seulement pour but de faire ressortir les dangers à éviter. Le lecteur ne devra pas en conclure que les études anciennes ont toutes connu chacun des défauts signalés.

On constatera aussi que les conceptions nouvelles sont encore tâtonnantes sur bien des points. L'appel à d'autres méthodologies que celle de la psychologie expérimentale est de nature à enrichir les moyens de recherche mis à la disposition des éducateurs. Cet élargissement, disons-le d'emblée, ne

suffit pourtant pas à doter le pédagogie scientifique d'une méthodologie cohérente. On verra en conclusion que des options fondamentales restent nécessaires.

I. Le but de l'évaluation

La pédagogie expérimentale est née de soucis très concrets. L'enseignement de l'orthographe atteint-il des résultats satisfaisants? Une autre façon d'aborder cet enseignement ne réussirait-elle pas mieux? L'apparition de méthodes concurrentes suscitait nécessairement des conflits: le recours à l'expérience devait permettre de trancher objectivement en faveur de l'une ou de l'autre. Une méthodologie simple en découlait. Il suffisait de comparer les résultats d'un groupe expérimental et d'un groupe de contrôle. Les progrès ultérieurs dans l'organisation des dispositifs expérimentaux ont contribué à affiner ces comparaisons; le principe d'une mise en compétition de méthodes différentes est ainsi resté fondamental dans la théorie classique de l'expérimentation pédagogique.

Lorsqu'on veut examiner cette conception de façon critique, la position de l'expérimentateur paraît forte à première vue, parce que parfaitement logique. Si une méthode a de meilleurs résultats qu'une autre, il faut la préférer. Qui pourrait le nier? – Il est intéressant, pourtant, de souligner combien cette façon d'aborder le problème restreint le domaine d'étude et rend fragiles les conclusions que l'on peut en tirer.

Il est sous-entendu, tout d'abord, que les acquisitions cognitives des élèves sont le critère essentiel sur lequel devra porter la comparaison. Que fait-on alors des autres objectifs éducatifs? On admet aussi implicitement que l'intérêt des autres parties engagées dans le processus éducationnel (parents, maîtres, autorités) s'efface entièrement devant l'impératif de l'apprentissage des élèves. Peut-on oublier les besoins des enseignants, les désirs des parents, les soucis des administrateurs, les pressions de l'opinion publique, et même, tout simplement, les goûts des enfants? L'expérimentateur avait réduit les dimensions du problème pour pouvoir le traiter plus facilement. Il reste à se demander si la réponse qu'il donnait alors avait encore un sens.

Cronbach (1964) fut l'un des premiers à critiquer cette conception de la recherche pédagogique et à demander un élargissement de son domaine d'étude. La mise en compétition de méthodes, par exemple, est incapable de donner des résultats généralisables, en raison de l'importance que prennent les facteurs du contexte, par rapport aux facteurs mêmes que l'on veut étudier (engagement affectif des maîtres ou des élèves, par exemple). Il est, de plus, dispendieux de recommencer des évaluations comparatives pour chaque nouvelle méthode ou à l'occasion de chaque nouvelle décision à prendre. Cronbach demande que l'on examine plutôt les résultats de l'enseignement par rapport à un large éventail d'objectifs éducatifs. On devrait connaître la performance des élèves d'abord dans chacune des dimensions pédagogiques et psychologiques visées par le nouvel enseignement. Par exemple, on devrait pousser l'enquête aux différents niveaux d'objectifs cognitifs, pour s'assurer de la capacité de généralisation des connaissances acquises; on devrait s'intéresser aux transformations d'attitudes résultant de l'enseignement. Par ailleurs, on devrait également contrôler des dimensions que le nouveau curriculum ne vise pas explicitement; pour s'assurer que le gain dans une direction n'est pas obtenu aux dépens d'une autre capacité aussi importante. Si l'on possède ces données pour chaque forme d'enseignement, on est en mesure de faire ensuite toutes les comparaisons utiles.

Ce premier élargissement fut le point de départ d'une évolution que Stake (animateur principal de la Conférence de Liège) a cherché à pousser à son terme extrême. Pour lui, poser le problème en termes d'objectifs éducatifs plus ou moins atteints est trop restrictif. La théorie des curriculums, et la pratique classique de l'évaluation, mettent en évidence la multiplicité des liens qui relie l'école aux divers secteurs de la société. Évaluer un nouveau curriculum implique, par conséquent, qu'on consulte tous les partenaires, qu'on parte des décisions à prendre par les responsables; qu'on détermine les

informations dont ils ont besoin ainsi que les questions que se posent les autres intéressés. Le chercheur qui s'astreint à cette enquête s'aperçoit de la multitude des dimensions à prendre en compte: économiques, sociales et politiques, juridiques et historiques, personnelles, éthiques, etc.

Dans cette perspective, la finalité de l'évaluation est d'abord pratique, au service de l'enseignement. Ainsi, l'évaluation élargie, appelée «*holistique*» par Wulf (1975), a pour but premier de fournir aux divers partenaires du système scolaire les informations dont ils ont besoin pour améliorer le fonctionnement de ce système. Ceci implique d'estimer, dans toute la mesure où c'est pratiquement réalisable, l'ensemble des facteurs qui paraissent significatifs aux intéressés, sans simplifier indûment la complexité des variables qui interviennent.

II. Le rôle des participants à l'évaluation

Le changement d'objectif auquel on vient de faire allusion explique le changement de rôle des partenaires de l'évaluation.

Dans l'optique traditionnelle, la responsabilité de toute l'entreprise d'évaluation appartient au chercheur. Expérimenter implique en effet un certain pouvoir sur les choses. Il faut pouvoir manipuler les variables pour en déterminer les effets. C'est ainsi seulement qu'on peut distinguer des relations causales de simples successions d'événements purement fortuites.

L'expérimentateur se donne d'emblée un rôle dirigeant. Même s'il doit négocier pour faire accepter sa vision des choses par les autorités scolaires, il obtient toujours l'accord des responsables sur un certain dispositif expérimental. C'est lui qui choisit souvent ce qu'il veut observer. C'est lui qui construit les instruments de mesure correspondants. C'est de ses propres statistiques, pratiquement invérifiables par d'autres, que dépend le verdict. Il est le «grand-prêtre»; il révèle la «Vérité», à laquelle, seul, il a accès. Sans doute les administrateurs choisissent-ils quelquefois une autre solution que celle qu'il leur «souffle». Ils ne peuvent le faire, cependant, sans enfreindre les lois de la rationalité que le chercheur représente. Lorsqu'on suit Cronbach, la position du chercheur est déjà moins impérieuse. On admet d'abord la multidimensionalité des effets de l'enseignement. Les méthodes ont alors peu de chances de se classer de la même façon sur toutes les dimensions étudiées. Le responsable scolaire peut privilégier d'autres objectifs pédagogiques que ne le fait le chercheur. Il peut ainsi choisir «rationnellement» une méthode pédagogique que le chercheur aurait, quant à lui, rejetée. Les jugements de valeur n'appartiennent plus aux scientifiques.

Pour Stake, l'attitude non-directive du chercheur doit aller plus loin encore. Les dimensions à évaluer en priorité doivent être déterminées par les intéressés. Ceux-ci peuvent dire aussi le genre de données qu'ils souhaiteraient recevoir. Sans doute le chercheur reste-t-il le spécialiste de l'information, celui qui peut faire valoir des considérations techniques, de fidélité de la mesure, par exemple. Son rôle cependant se limite à celui de «pouvoyeur d'information valide». En tant que chercheur, il n'a pas à se prononcer sur la démarche que suit un responsable pour parvenir à sa décision. Il doit seulement fournir les éléments d'information dont ce responsable a besoin.

Mieux encore, l'évaluateur se situe généralement au sein d'un réseau de relations sociales. Les élèves, leurs parents, les enseignants, les commissions scolaires, les autorités cantonales, et bien d'autres groupes encore ont leur point de vue particulier, ont des opinions à défendre. Il est exclu, dans la plupart des cas, qu'un jugement unanime ressorte de ces confrontations. Le chercheur n'a pas à guider ces positions diverses vers un accord. Il doit seulement faciliter la communication entre ces groupes, expliciter «objectivement» leurs points de vue «subjectifs». Il peut ainsi être amené à présenter des interprétations contradictoires de la même réalité, sans qu'on puisse lui demander de trancher dans un sens ou dans un autre. Il reflète simplement aux uns et aux autres leurs perceptions respectives. Cette décentration du chercheur se situe dans la ligne de la réflexion de certains psychologues,

comme Argyris (1970), qui ont examiné à quelles conditions un consultant pouvait intervenir dans une organisation sans créer du même coup un problème de dépendance à son égard. Les anthropologues sont parvenus par une autre voie aux mêmes conclusions, en définissant leur rôle comme celui d'«observateurs participants». La distance est grande entre cette conception du rôle du chercheur (et du but de son intervention) et celle de l'expérimentateur traditionnel, supposé objectif et extérieur au système qu'il étudiait. Au départ, le chercheur s'appropriait l'évaluation. A l'arrivée, ce n'est plus l'«évaluateur» qui évalue. Il s'en défend, au contraire, pour mieux laisser s'exprimer les différents points de vue en présence et conserver aux responsables leur pleine liberté de décision.

Le rôle des autres membres du système scolaire a naturellement évolué en sens inverse. De simples spectateurs passifs, ils sont devenus participants dans l'entreprise commune. Leur point de vue est sollicité. On leur rapporte le point de vue des autres. L'évaluateur élargit leur champ d'appréhension du réel, en mettant à leur disposition des informations qu'ils n'auraient pas pu acquérir par eux-mêmes. C'est le cas des données économiques, des résultats statistiques, des avis juridiques, des documents historiques, des expertises et enquêtes, etc., qu'il leur fournit. Munis de cet ensemble de points de vue, apercevant la réalité sous tous ses éclairages, ils peuvent prendre leurs décisions plus librement, parce que moins conditionnés par leur seule expérience personnelle. Enfin, ils risquent moins de devoir faire face à des situations imprévues, parce qu'ils maintiennent l'ensemble de l'expérience «sous contrôle permanent».

III. L'objet de l'évaluation

Vouloir, comme on l'a dit plus haut, estimer l'ensemble des facteurs qui paraissent significatifs aux intéressés implique d'élargir l'objet de l'évaluation. On vient également de faire allusion à la diversité des éclairages nécessaires à une prise de décision équilibrée. Quelles sont les variables que l'évaluateur doit considérer?

1. Les effets

Au départ, dans les études d'évaluation classiques, les seules variables que les chercheurs prenaient en considération étaient les progrès des élèves. Les objectifs éducatifs une fois définis, ils s'efforçaient de quantifier ces progrès au moyen de tests pédagogiques appropriés. Dans certains cas, ils complétaient l'étude des effets cognitifs du curriculum par celle de ses effets effectifs. Cronbach avait demandé que l'on étudie aussi les effets de la formation sur certaines dimensions de la personnalité qui n'étaient pas visées par les auteurs de méthodes. Ce contrôle, beaucoup trop lourd pratiquement, fut rarement effectué.

Il suffit de penser à l'exemple des cours programmés du type «Skinner» pour constater que l'évaluation pédagogique, telle que la pratiquent les divers sous-groupes du système scolaire ne suit pas ce modèle théorique. L'effet de ces cours sur l'apprentissage des élèves est bon: d'innombrables expériences l'ont confirmé. Pourtant leur emploi se développe peu. C'est que leur effet sur la motivation des élèves est plutôt négatif à terme, à cause de leur monotonie. C'est aussi que le corps enseignant se sent menacé dans son rôle traditionnel et qu'il réagit négativement à leur introduction. Autrement dit, l'effet attendu sur l'apprentissage n'est pas le seul critère auquel on se réfère et les élèves eux-même ne sont pas le seul élément du système scolaire que l'on prenne en considération.

Il importe aussi d'étudier les conséquences d'une nouveauté pédagogique sur la formation exigée du corps enseignant, sur le travail qu'il doit fournir en dehors de l'école, sur les possibilités de relations maître-élèves et élèves-élèves, sur le «curriculum caché», c'est-à-dire les valeurs implicites transmises par l'école, etc. Tous ces aspects qualitatifs pèsent probablement plus lourd dans l'acceptation ou le rejet d'une réforme que quelques points de plus ou de moins à un test de connaissances.

Les responsables de l'école et les autorités politiques s'attachent probablement davantage à des critères sociologiques, économiques, juridiques... Les parents sont sensibles aux répercussions du nouveau curriculum sur la vie de famille, etc. Toutes ces réactions sont compréhensibles, légitimes et méritent d'être prises en considération.

Il importe de se rendre compte que pour rassembler des informations valides dans tous ces domaines une équipe pluridisciplinaire est nécessaire. Et pourtant, aussi longue que puisse être l'énumération esquissée ci-dessus, la prise en compte de tout l'éventail des effets possibles ne suffit pas encore à décrire l'objet de l'évaluation élargie. Il faut en effet apprécier les conditions dans lesquelles ces résultats ont été obtenus.

2. Les conditions

Contrairement à la position prise par certains évaluateurs américains, comme Scriven dans un texte préparé pour la Conférence de Liège, la comparaison brutale des résultats obtenus par divers curriculums ne traduit pas leur efficacité véritable. Sans doute le rendement observé est-il bien réel; mais la généralisabilité des conclusions peut être mise en doute, si l'on ne sait pas dans quelles conditions particulières ces résultats ont été obtenus.

Ainsi, l'enthousiasme ou le don pédagogique de tel enseignant peuvent expliquer le succès d'une méthode, alors que les mauvaises relations maîtres-élèves ou simplement le fait d'avoir gardé la classe après l'heure pour terminer le test peuvent expliquer des résultats défavorables. Donald Campbell (1974) a insisté récemment sur la priorité à donner à ces informations qualitatives dont dépend finalement la validité des résultats quantitatifs et des interprétations qu'on en tire.

Une évaluation qui vise à fonder des décisions judicieuses ne peut donc négliger l'analyse approfondie de tous les aspects du système: le contexte socio-culturel des élèves, le niveau de compétence des enseignants, le climat qui a entouré le lancement de la réforme, la signification qu'elle a prise pour les divers sous-groupes impliqués dans l'expérience, etc.

3. Les processus

Expliquer les résultats par les conditions dans lesquelles ils ont été obtenus, c'est dire «comment on en est arrivé là». On voit que, dès que l'on veut interpréter un résultat brut, on commence à introduire des variables d'un autre ordre et à se rapprocher du type d'explication historique. Par exemple, on passe du domaine «objectif» au domaine «subjectif» lorsqu'on fait appel à la signification que prend telle condition objective pour les personnes impliquées dans l'étude. La transition est ainsi insensible de l'étude des conditions d'une expérience à la description de son déroulement historique. Effectivement, c'est bien introduire un certain type de compréhension causale que de dire comment quelque chose s'est produit.

L'étude d'évaluation peut ainsi prendre la forme d'une «étude de cas». Sa publication permet de mettre à la disposition d'autres personnes et d'autres publics un «résultat d'expérience». Pour être de type historique, cette connaissance n'en est pas moins pertinente et susceptible d'un certain degré de généralisation.

4. Pourquoi même les effets?

L'étude des processus peut naturellement se poursuivre à d'autres niveaux, individuels en particulier. On ne peut porter un jugement valable sur une méthode d'enseignement qu'après avoir suivi dans le détail le processus de pensée qu'elle suscite chez l'élève. Il ne suffit pas en effet d'obtenir de lui la bonne réponse dans une condition standardisée. L'enfant doit acquérir une compréhension qui lui permette aussi de résoudre d'autres problèmes du même type. L'emploi de la méthode clinique piagétienne semble le moyen le plus approprié pour examiner le degré d'assimilation des connaissances acquises par l'enfant.

On en vient aujourd'hui à privilégier l'étude des processus d'acquisition de la connaissance. Partant de conceptions très différentes, J. Brunner (1973) et R. Glaser (1973) arrivent à la même conclusion: la façon dont l'enfant apprend est plus importante que ce qu'il apprend. Son mode d'acquisition prépare en effet ses méthodes d'apprentissage ultérieures. Or c'est dans ce domaine que se situe l'essentiel de la formation que peut donner l'école. Les performances aux tests de rendement perdent ainsi de leur intérêt, par opposition aux démarches par lesquelles sont acquises ces connaissances, dans la pédagogie que l'on expérimente. Ce déplacement de leur objet d'étude éloigne toujours plus les études d'évaluation de leur modèle expérimental originel.

IV. Les instruments de l'évaluation

La prise en compte de tout cet ensemble de variables nouvelles n'a de sens que si l'on peut trouver des moyens d'observation pour chacune d'elles. Nous allons mentionner ci-dessous les techniques à disposition.

On souhaiterait, pour des raisons pratiques, employer des instruments simples, rapides, mais cependant précis. Ce n'est pas toujours possible. C'est pourquoi des compromis sont nécessaires, comme en radio, entre «largeur de bande» et «précision». Des choix s'imposent également, telle ou telle dimension prenant plus d'importance dans tel contexte que dans tel autre.

La liste proposée ci-dessous n'implique donc pas que chaque étude d'évaluation doive utiliser tous ces types d'instruments. Nous voulons simplement mettre en évidence l'élargissement de l'arsenal habituel de la pédagogie. Surtout, nous pensons utile de reproduire à cette occasion les commentaires de Wulf (1975) sur les précautions à prendre lorsqu'on emploie chacun de ces instruments. Wulf considère que les partenaires essentiels de l'évaluation décentralisée sont les chercheurs et les enseignants. C'est pour cette raison qu'il présente séparément les instruments favorisés des uns et des autres:

1. Méthodes d'évaluation à la disposition des chercheurs¹

1.1 Méthode expérimentale et quasi-expérimentale

(S'appuyant sur la prise de tests, le relevé d'indicateurs divers, comme le taux d'absentéisme, le nombre de punitions, etc.)

Conditions: L'ensemble du dispositif expérimental doit avoir été mis au point conjointement par toutes les parties concernées, dans l'esprit d'une recherche-action.

Problèmes à prévoir: Dans quelle mesure les résultats sont-ils affectés par la conscience que prennent alors les participants d'être les sujets de l'expérience? Le plan expérimental contrôle-t-il cet effet? Existe-t-il pour le chercheur un conflit de rôle entre sa situation de collaborateur et d'évaluateur? Est-il à la fois juge et partie?

1.2 Emploi de questionnaires standardisés, d'entretiens structurés, de fiches d'observation, de relevés d'emploi du temps, de listes d'incidents critiques, etc.

Conditions: Leur but doit être expliqué et l'on doit au moins discuter après coup les résultats avec les intéressés.

Problèmes à prévoir: Aux conflits de rôle déjà mentionnés, s'ajoute le danger, pour le chercheur, de présenter ses résultats d'une façon trop définitive, qui ne permette pas aux autres participants de contribuer à leur interprétation.

1.3 Analyse de contenu

Problèmes: Cette procédure s'appuie sur des documents (programmes, manuels, journaux de classe

¹ Une autre liste structurée décrivant les techniques de collecte de données est présentée par De Bruyne Herman et De Schoutheete (1975, p. 202 - 205), avec également l'indication des obstacles à prévoir et des avantages relatifs de chacune.

p. ex). On ne sait pas dans quelle mesure la réalité (p. ex. l'enseignement réellement donnée) s'écarte des indications mentionnées.

1.4 *Interview ouverte*

Conditions: Il faut au moins discuter les conclusions avec les intéressés. Dans la mesure du possible, les questions à poser devraient être mises au point en commun au préalable.

Problèmes: Comme sous 1.2, la discussion des conclusions risque d'être escamotée.

1.5 *Observation participante*

Conditions: Le chercheur qui assure l'observation doit se considérer réellement comme un collaborateur pleinement engagé.

Problèmes: La position d'observateur conduit le chercheur à prendre un certain recul, qui peut avoir un effet négatif sur les relations entre lui-même et les membres de son groupe.

1.6 *Compte-rendus subjectifs (protocoles, lettres, rapports, etc.)*

Conditions: Il faudrait, si possible, systématiser la façon d'établir ces rapports.

Problèmes: Sinon, l'évaluation d'impressions assez hétérogènes parce que choisies au hasard, risque d'être très difficile.

1.7 *Contacts suivis avec des scientifiques ou des spécialistes (par lettres, entretiens, séminaires)*

Conditions: Il faut pouvoir justifier le choix de ces personnes (par leur compétence, leur activité, etc.)

1.8 *Jeux de rôles, simulations, pour anticiper ou détecter des difficultés éventuelles*

Conditions: La situation du jeu doit rester dans le cadre des possibilités attendues.

1.9 *Comparaison avec la littérature*

Problèmes: Les facteurs que l'on mentionne dans la littérature ne s'appliquent pas nécessairement au projet en cours.

1.10 *Informations imprévues (articles de journaux, lettres, etc.)*

Conditions: Le recueil de ces documents doit se faire de façon systématique. (Il ne faut pas conserver seulement les articles élogieux, ou critiques, p.ex.).

2. Méthode d'évaluation à la disposition des enseignants

2.1 *Tests informels*

Conditions: Les élèves doivent au moins participer à la correction. La procédure d'évaluation doit devenir claire pour eux.

Problèmes: On peut surévaluer l'importance des tests (bachotage). De plus les tests expriment des objectifs éducatifs sur lesquels ultérieurement l'enseignement tend à s'aligner. Si l'on ne mesure que les connaissances, ce type d'objectif tendra à recevoir alors un poids excessif.

2.2 *Sociogrammes*

Conditions: Le maître ne doit pas les utiliser à ses propres fins d'enseignement. Il doit discuter les résultats avec les élèves. Toutes les dispositions à prendre à leur propos doivent être prévues en commun.

2.3 *Questionnaires informels, procédures d'observation, etc. (du genre des tests tirés de magazines)*

Conditions: Les élèves doivent au moins prendre part à l'évaluation.

Problèmes: Ces instruments ne répondent pas aux exigences de tests véritables et ne doivent pas être pris trop au sérieux. On risque alors de les considérer seulement comme des jeux.

2.4 *Listes de problèmes, à pointer par les élèves*

Conditions: Les élèves devraient participer dès le début à l'établissement du catalogue des problèmes. Le traitement des résultats devrait se faire aussi en commun.

2.5 *Comptes rendus personnels*

Conditions: Les élèves ne doivent pas craindre de répercussions négatives s'ils formulent des critiques.

Problèmes: Faute de base de comparaison, on risque de privilégier certains aspects dans le résumé qui sera fait de ces rapports.

2.6 *Évaluation réciproque d'enseignants travaillant en équipe*

Conditions: Les relations entre ces maîtres doivent leur permettre d'accepter la critique, sans que des conflits personnels en résultent.

Problèmes: L'isolement typique de l'enseignant représente une difficulté majeure.

2.7 *Discussion des exigences (remise en question continue des normes de l'enseignement, avec les élèves ou des collègues)*

Conditions: Une certaine habitude de ce genre de discussion est nécessaire.

Problèmes: Plus on s'identifie au plan d'étude, plus il est difficile de prendre le recul nécessaire pour le remettre en question.

2.8 *Techniques d'animation de groupe*

Conditions: L'enseignant ne doit utiliser que les techniques qui ne nécessitent pas de formation particulière.

Problèmes: L'évaluation ne doit pas dégénérer en dynamique de groupe.

2.9 *Répétition de travaux (portant sur des connaissances, des comportements, etc.)*

Conditions: Il faut protocoler la façon dont le problème a été traité la première fois, pour qu'on puisse établir une comparaison, par exemple entre le début et la fin d'une séquence d'apprentissage.

2.10 *Informations imprévues (commentaires des parents, critiques des élèves, etc.)*

Conditions: Procédure de recueil systématique.

Ces divers instruments restent encore surtout du domaine de la psychologie ou de la sociologie. Huberman (1973) cite d'autres méthodes utilisées pour évaluer l'efficacité d'un système scolaire. Certaines sont de nature économique, font référence aux possibilités d'emploi ultérieures des élèves, d'autres résultent des études de pédagogie comparée, comme l'indice de différenciation structurale du système d'un pays. Le recours à des spécialistes de l'analyse des systèmes permet d'autres types d'évaluation encore, comme le montrent les études de Haag (1975) et Munari (1975). Il ne fait pas de doute que pratiquement toutes les sciences qui traitent de l'homme d'une façon ou de l'autre peuvent apporter leur contribution à l'évaluation des expériences pédagogiques.

Le tableau présenté par Smith et Pohland (1974) peut servir de conclusion, parce qu'il situe bien le rôle respectif des trois disciplines les plus fréquemment impliquées dans l'évaluation, et parce qu'il évoque les principaux instruments d'observation qui leur correspondent (voir Tableau 1).

Selon ces derniers auteurs, c'est l'anthropologie qui fournit le modèle méthodologique le plus pertinent pour l'évaluation élargie. Ceci implique que le point de vue subjectif des participants et la vision propre de leur situation soient pris en compte et acceptés comme objets d'étude valables. On peut même considérer ces «représentations», ces «significations», ces «symboles» comme variables explicatives, au même titre que les paramètres objectifs de la situation d'enseignement. On assiste ainsi à un retournement de perspective. La pédagogie se rapproche des sciences humaines, comme l'histoire, et s'éloigne des sciences expérimentales, comme la biologie.

Il reste à examiner maintenant les conséquences de ce retournement sur la méthodologie de l'évaluation.

Tableau 1: Méthodologie de l'évaluation pour l'étude de cas de Smith et Pohland (1974)

Base théorique et technique	Procédé de recherche	Instruments	Rôles	Sujets d'étude
1. Psychologie	Dispositif expérimental Groupes expérimentaux et contrôle placés au hasard dans les classes Tests de rendement avant et après expérience	Tests de rendement scolaire standardisés	Evaluation sommative	Elèves
2. Sociologie	Enquêtes d'opinion Construction d'indices de préférence	Questionnaires d'attitudes structurés pour maîtres et élèves Interviews de parents	Evaluation sommative Maîtres	Elèves Parents
3. Anthropologie	Observation participante Etude sur le terrain	Observateurs: notes prises sur-le-champ, entretiens libres, documents	Evaluation formative et sommative	Elèves Maîtres Matériels Buts

V. Les méthodes de l'évaluation

Comme plus haut, les tendances nouvelles apparaîtront plus clairement si on les oppose point par point à la méthodologie plus ancienne que Stake dénomme l'évaluation «quantificatrice». Il n'existe pas, bien sûr, de méthodes aussi typées dans la réalité. Dans les documents qu'il a distribués à Liège, Stake présentait 9 types de méthodes et pensait encore pécher par excès de simplification. Les positions extrêmes, pourtant, sont plus faciles à distinguer.

1. Désignations

L'évaluation quantificatrice est appelée ainsi parce qu'elle recueille en priorité des données chiffrées, alors que les données subjectives, les attitudes et les jugements par exemple, ne sont recueillis qu'à titre de reflets indirects de ces effets objectifs. L'accent est mis sur les résultats, qui doivent pouvoir se mesurer.

La position opposée, que nous appellerons l'évaluation qualificatrice, recueille au contraire des descriptions plus qualitatives. Les résultats statistiques ou chiffrés peuvent être présentés également, mais leur signification est considérée comme finalement moins riche. Les perceptions et les représen-

tations des participants à l'expérience constituent l'élément de base de l'évaluation. L'accent est mis sur les processus, sur les événements qui ont conduit aux résultats.

2. Cadres conceptuels respectifs

Derrière ces deux modes d'approche de l'évaluation, on peut distinguer deux conceptions différentes de la causalité, en explication et en compréhension.

Pour l'évaluation quantificatrice la causalité est une relation observée de l'extérieur entre trois ordres de faits. Il existe des variables *indépendantes* dont on peut étudier l'effet sur des variables *dépendantes*, à condition de *contrôler* l'effet de l'ensemble des autres variables. Seront variables indépendantes, celles que l'expérimentateur modifiera à sa guise (en général le curriculum étudié). Seront variables dépendantes, celles qui révéleront l'effet du curriculum (les dimensions d'objectifs pédagogiques). Les autres influences seront contrôlées en partie par des plans expérimentaux ingénieux, mais surtout par l'attribution au hasard des élèves et des maîtres aux diverses conditions et par les tests statistiques ultérieurs.

Pour l'évaluation qualificatrice, la causalité est plus une relation entre des significations. Ce sont les représentations que les différents partenaires se font de la réforme et leurs perceptions réciproques qui déterminent leurs actions. Le chercheur dégage un genre de causalité compréhensive, une totalité signifiante, comme le dit De Bruyne (1975, p. 160).

La mise en évidence de ces relations suppose, d'après Weiss et Rein (1970), que l'on détermine trois ensembles de données qualitatives: les limites du *système* à étudier, la nature des *problèmes* soulevés, et le déroulement des *événements*, voire des affrontements, relatifs à ces problèmes. Les cadres conceptuels correspondants relèvent naturellement de l'analyse des systèmes, de la politologie et de l'histoire. L'apport de l'analyse des systèmes est d'abord de guider la recherche d'un point d'équilibre entre le désir de tout observer et le besoin de limiter le travail. Ce point de vue suggère de concentrer son attention sur le plus petit ensemble d'individus ou de groupes qui peuvent rendre compte de ce qui se passe, c'est-à-dire décrire et expliquer la suite des événements à l'intérieur du système. L'analyse des systèmes donne ensuite un principe clair pour délimiter le domaine à décrire: c'est l'ensemble des éléments qui sont en interaction à propos du problème étudié. Cet ensemble varie selon le problème; il varie aussi dans le temps. Si certaines forces affectent le système, mais ont peu d'interaction avec lui (une crise économique par exemple), il suffit de les traiter comme variables exogènes: on n'a pas à les analyser pour elles-mêmes. Il faut étudier, par contre, des forces extérieures, si elles sont affectées en retour par les variables relatives à l'école. Si une réforme devient un enjeu politique, par exemple, les prises de position des partis deviennent des variables internes au système, et donc à inclure dans l'étude d'évaluation.

La politologie offre des outils conceptuels et des méthodes pour décrire le problème. Elle dirige l'attention sur les conflits et les coalitions. Elle découvre les groupes d'intérêts en présence. Elle analyse leurs actions en termes de stratégies plus générales. Elle explique par là leurs réactions particulières au programme de réforme proposé.

L'histoire enfin a une méthodologie propre utilisable pour le recueil des témoignages, leur confrontation et leur évaluation (critique des sources), ainsi que pour la description et l'interprétation du développement des événements. Elle part de trois types de données de base: observation des événements, compte rendu de témoins, analyse de documents.

Ces sources d'information doivent être comparées et combinées pour que l'on puisse parvenir à une vue d'ensemble cohérente et significative du déroulement de l'expérience pédagogique étudiée.

3. Relation au temps

En principe, l'évaluation quantificatrice est indépendante du temps. Une première raison est le caractère immuable d'un dispositif expérimental. La nature des hypothèses détermine les procédures à

suivre pour le recueil des données. On a pu dire que, dans l'idéal (de la méthode hypothético-déductive), il devrait être possible d'écrire un rapport scientifique avant même d'avoir fait la moindre observation. Les hypothèses de travail devraient être suffisamment bien précisées et la déduction à partir de ces hypothèses devrait être assez rigoureuse pour que l'expérience n'ait plus qu'à apporter la confirmation attendue. Il s'ensuit qu'un plan expérimental est rigide. Il faut donc au chercheur une situation stable, où les choses se passent, jusqu'à la fin de l'étude, de la façon dont le chercheur les avait prévues.

Une seconde raison tient aux techniques d'analyse. Comme il lui faut l'ensemble des données pour effectuer ses tests statistiques, le chercheur est amené à prendre en charge de préférence des évaluations sommatives. De tels bilans sont essentiellement statiques, parce que définitifs. Ils sont donc extratemporels.

L'évaluation qualificatrice, au contraire, se situe d'emblée dans le temps. L'observateur suit ce qui se passe. Sa méthode est fondamentalement souple. Il réagit à ce qu'il perçoit, ou à ce qu'on lui demande. Stake parle d'une évaluation «répondante»; on pourrait aussi dire «interactive». Santini (1975), par exemple, montre comment les divers groupes concernés par une innovation pédagogique, à Zoug, ont pu déterminer avec le chercheur les objectifs qu'il leur paraissait important d'évaluer. Pour l'étape suivante de la formulation des évaluations, l'interaction chercheur-participants fut tout aussi active. Pour Santini, comme pour Wulf, l'observateur-participant, qui vit avec le groupe qu'il étudie, peut prendre conscience des problèmes dès qu'ils apparaissent et en avertir ses partenaires dans l'expérience. Il peut ainsi prendre en charge une évaluation formative, c'est-à-dire une évaluation qui permet d'intervenir à temps et de faire réussir l'expérience, plutôt que de la juger.

Le chercheur qui intervient sur le cours d'une réforme peut-il même encore la juger? La réponse n'est pas simple, car elle fait appel à des conceptions différentes de la validité du jugement scientifique, selon que l'on considère l'un ou l'autre des deux types opposés d'évaluation.

4. Généralisabilité et validité

L'évaluation quantificatrice fonde ses généralisations sur la répétabilité des résultats observés. Les tests statistiques ont pour but d'assurer le chercheur de la stabilité de ses conclusions d'un échantillon à un autre. La valeur de ses mesures est assurée par des mises en relation préalables (études de fidélité) qui contrôlent également leur caractère répétable. La généralisabilité d'une relation observée se fonde dans ce cas sur la même démarche que l'inférence statistique, par laquelle on estime un paramètre d'une distribution à partir d'une série d'observations tirées au hasard dans cette population. Dans la réalité sociale globale où se situent les innovations pédagogiques, les situations ne sont pourtant ni contrôlables, ni répétables, comme le notent Weiss et Rein (1970) en relevant les difficultés théoriques et pratiques que soulève l'approche quantificatrice. C'est pourquoi l'évaluation qualificatrice s'est dégagée de la conception aristotélicienne de classes de situations équivalentes. Chaque cas est réellement considéré comme unique. Dans la mesure pourtant où l'on peut dire comment on est parvenu à l'état de fait actuel, la compréhension de cette évolution passée peut permettre d'effectuer des généralisations à d'autres situations futures. Les étapes du processus découlent logiquement de la finalité poursuivie. On se fonde alors sur des lois d'un autre type, qui explicitent la rationalité que l'action développe, qui décrivent la dynamique de l'articulation «buts-moyens» (cf. De Bruyne, 1975, p. 138).

Quant à la valeur des observations recueillies, elle peut être contrôlée, dans l'évaluation qualificatrice, par l'accord des divers observateurs. L'intersubjectivité tient lieu d'objectivité dans le domaine des relations humaines. Smith et Pohland (1974) présentent le schéma d'une matrice multiméthodes, multipersonnes, multisituations et multivariées. Une analyse de variance pourrait nous renseigner sur les composantes de variance dues à chacune de ces sources de variation, s'il s'agissait de données

quantitatives. L'accord (qualitatif) entre les points de vue exprimés par divers groupes peut être considéré comme une validation logiquement analogue.

On arrive ainsi à une interprétation nouvelle de la validité d'une étude d'évaluation, telle que la conçoit Wulf (1975) à la suite d'autres auteurs. Cette validité se caractérise, selon lui, par les quatre critères suivants:

- a) *transparence*: reproductibilité du processus de jugement pour tous les partenaires, sur la base d'une explicitation des fonctions, des intentions, des rôles et des méthodes de l'évaluation,
- b) *cohérence*: accord des procédures utilisées avec les intentions annoncées,
- c) *acceptabilité*: reconnaissance par les participants du caractère indiscutable des résultats présentés,
- d) *pertinence*: importance des résultats de l'évaluation pour la prise de décision envisagée.

On remarque combien l'accent est mis dans ce cas sur une saisie significative de l'expérience pédagogique étudiée dans ce qu'elle a de singulier. On recherche l'accord intersubjectif pour mieux fonder une décision qui est elle-même unique. Cette conception s'oppose à l'intention généralisatrice des sciences naturelles qui néglige tout ce que l'expérience a de singulier et cherche au contraire à fonder des stratégies d'ensemble.

5. Autres méthodes utilisées

Smith et Pohland notent que, même au sein des études de type anthropologique, les styles de recherche se différencient. Ils notent quatre dimensions principales de variation: importance accordée à la description des événements, à la génération de théories, à la vérification de lois et à la quantification. Dans un document présenté à Liège, Stake propose de son côté huit dimensions descriptives qui permettraient d'établir une classification plus fine des modèles d'évaluation. Il donne des exemples d'évaluations réalisées au moyen d'autoévaluations, de discussions en commissions, d'études sur dossiers, d'analyses de gestion, d'explicitations de politique sociale, et même de conférences contradictoires.

C'est dire qu'il n'existe pas une méthode nouvelle mais un ensemble de recherches dont le caractère commun est l'ouverture à de nouveaux critères et l'élargissement du schéma expérimental originel. Le choix entre ces variantes est affaire d'opportunité. Il dépend surtout de l'importance qu'accordent les demandeurs aux divers types d'information.

Ce qui caractérise enfin les recherches d'évaluation actuelles, c'est justement l'intérêt pour la méthodologie pour elle-même. Alors que le modèle scientifique de l'évaluation comparative avait été considéré plus ou moins comme acquis et intangible pendant de nombreuses années, il est maintenant remis en cause. Chaque étude de cas sert à expérimenter, en même temps qu'une réforme particulière, une nouvelle façon d'aborder le problème de l'évaluation en général.

VI. Les étapes de l'évaluation

Caractériser abstraitement une méthodologie nouvelle par opposition à une autre, comme on vient de le faire, ne suffit pas à guider le chercheur dans sa démarche concrète. C'est pourquoi les lignes ci-dessous vont s'attacher à détailler les problèmes qui se posent à chaque étape d'une entreprise d'évaluation. On suivra le schéma d'ensemble proposé par Parlett (1974) sous le nom d'évaluation «illuminative».

1. La mise en train de l'évaluation

Il serait contraire à la conception adaptative de l'évaluation élargie de la figer dans une façon de procéder rigide, applicable dans toutes les situations. Au contraire, elle ne constitue qu'une stratégie générale: un ensemble différent de techniques doit être mis en œuvre dans chaque cas.

Cette nécessité de répondre au besoin ne fait que rendre plus importante la première étape, celle justement où le but à atteindre doit être déterminé. Les questions suivantes paraissent fournir une base utile pour l'analyse du problème.

- a) Quel est le problème principal qui suscite le besoin d'une étude d'évaluation?
- b) Quels sont les autres problèmes, plus ou moins liés au premier, qui motivent (explicitement ou non) la recherche envisagée?
- c) Quelles sont les décisions à prendre?
- d) Quel en est l'enjeu (pédagogique, économique, politique, etc.)?
- e) Quelles informations seraient utiles pour fonder ces décisions?
- f) A quoi veut-on arriver (pilotage ou bilan de la réforme)?
- g) Quelle (s) question(s) pose-t-on à l'évaluateur?
- h) Quelles stratégies de recherche découlent des questions (implicites et explicites)?
- i) Quels sont les autres personnes, groupes, publics intéressés à l'évaluation?
- j) Quelles sont leurs motivations à la recherche?
- k) Quel est l'enjeu pour eux?
- l) Participent-ils à la discussion, influencent-ils les décisions?
- m) De quelles informations ont-ils besoin?
- n) Recevront-ils le(s) même(s) rapport(s) que l'autorité qui commande?
- o) A quelles informations l'évaluateur aura-t-il accès?
- p) Quelle diffusion pourra-t-il donner au(x) rapport(s) qu'il fera?
- q) Quel rôle l'évaluateur est-il censé jouer vis-à-vis de son mandant et des divers autres publics?
- r) Quels rôles risque-t-il de jouer, s'il n'y prend garde, vis-à-vis d'eux?
- s) Comment risque-t-il d'être perçu par chacun de ces milieux?
- t) Quel effet cela peut-il avoir sur les informations qu'il recueillera?
- u) Que peut-on faire pour augmenter la validité des informations qui lui seront fournies?
- v) Que peut-on faire pour assurer la validité des conclusions qu'il fournira, c'est-à-dire pour assurer la transparence, la cohérence, l'acceptabilité et la pertinence de sa procédure?
- w) Quel type d'étude en découle – et quel type d'évaluateur?
- x) Quels seront le statut et la marge de manoeuvre de l'évaluateur?
- y) Dans quelles limites (hiérarchiques, territoriales, temporelles, budgétaires...) devra-t-il se tenir?
- z) Par quels canaux pourra-t-il communiquer avec tous les intéressés?

Cette étude préalable nécessitera déjà des contacts multiples avec tous les milieux concernés, et non pas seulement avec les mandants. Dans certains cas, en effet, l'évaluation demandée peut prendre une signification polémique et le chercheur peut se trouver entraîné dans un conflit qui rendrait ses conclusions d'avance inacceptables.

Il faudra aussi que, par des moyens d'information appropriés, tous les milieux avec lesquels il devra prendre contact soient persuadés qu'il ne vient pas les inspecter ou les remettre en cause, mais qu'il doit plutôt étudier avec eux le fonctionnement du système dont ils font partie.

Cette première analyse devra être assez approfondie pour poser et délimiter le problème. Elle devra se terminer par un contrat, spécifiant la demande. Elle ne devra pas, pourtant, engager l'avenir en limitant d'avance les domaines d'investigation possibles, les instruments à utiliser, etc. Elle ne doit fermer aucune porte. Elle doit seulement préparer la seconde étape, en ouvrant la voie à une exploration plus systématique.

2. L'exploration ouverte

Cette seconde phase est d'ordinaire la plus longue et la plus importante. Le chercheur doit se familiariser avec la réalité journalière de l'école. Il doit passer beaucoup de temps à visiter les écoles impliquées dans l'étude, à observer ce qui s'y passe, à écouter maîtres et élèves, à découvrir les caractéris-

tiques particulières de chaque école. Il doit se comporter comme un anthropologue qui examine une culture différente de la sienne et essaye de la comprendre «de l'intérieur».

Durant cette phase, le chercheur s'informe sur tous les aspects de la réforme et recueille une masse d'informations aussi diverses que possible. Il examine le comportement des maîtres et des élèves, leur accord avec les intentions annoncées du plan d'études, la cohérence entre ces principes et les mesures administratives prises dans les écoles réformées ou les autres écoles, antérieures et postérieures. Il évalue les coûts, apparents et cachés. Il analyse les tâches données aux élèves pour déterminer les objectifs éducatifs implicites. Il décrit le «curriculum caché» qui ressort de l'organisation même de l'institution scolaire. Ses enquêtes ne se limitent pas d'ailleurs aux écoles. Il recherche les fondements théoriques de la réforme étudiée. Il essaye d'établir son histoire et de retracer son évolution. Il cherche à connaître l'attitude vis-à-vis de cette innovation des divers groupes sociaux qui y sont associés. Il observe les relations de ces groupes entre eux. Il essaye de sentir les coalitions qui se forment et les lignes de force ou les tensions sous-jacentes.

En bref, il essaye de se faire une idée aussi complète et diversifiée que possible de tous les aspects de la réforme, et de leurs multiples prolongements. Il le fait en essayant de partager l'expérience de tous les groupes concernés. Pour aller plus loin que cette première impression, il devra alors aborder la phase 3.

3. L'étude concentrée

La troisième phase commence lorsque certaines hypothèses ont pu être dégagées, sur la base de similitudes ou de différences observées, de régularités ou d'incidents critiques, etc. Ces hypothèses deviennent alors l'objet d'enquêtes plus poussées, destinées à les étayer ou à les infirmer. Entretiens, questionnaires, épreuves papier-crayon, observations systématiques, analyses de textes écrits, de documents, tout peut servir à approfondir et à valider l'information pressentie à l'étape précédente. Aucun instrument n'est exclu, mais aucun n'est privilégié non plus; les tests de connaissances ne représentent dans cette optique qu'un des aspects de la réalité à étudier.

La difficulté est de concentrer l'étude progressivement, de façon, bien sûr, à ne pas passer à côté de problèmes importants qui n'étaient pas suspectés au départ, mais de façon en même temps à ne pas accumuler une masse d'informations dont l'analyse serait difficile ultérieurement. La réduction des données se fait par étapes, le rôle du chercheur étant de détecter et d'identifier de façon de plus en plus précise les domaines qui font problème, ou qui représentent au contraire des progrès substantiels.

4. L'effort d'interprétation

Le chercheur peut alors commencer à organiser ces informations. C'est à ce moment qu'il pourra faire appel à diverses approches d'utilité générale dans les sciences humaines, en utilisant par exemple les concepts de l'analyse des systèmes, en recherchant des explications fonctionnelles, en établissant des similitudes structurales, en comparant des groupes, en construisant un modèle, etc.

En effet, le chercheur doit faire plus que de restituer ce qu'il a entendu, même sous une forme détaillée ou vivante. Il doit organiser sa description, y introduire un ordre fondé si possible sur une interprétation ou une explication. Assez souvent, il pourra se laisser guider par des questions très générales du genre des suivantes:

- a) Quels étaient les objectifs officiels de la réforme?
- b) Quelles étaient les intentions implicites des divers partenaires?
- c) Quels étaient, pour chacun, les problèmes justifiant une évaluation?
- d) Quels en étaient les implications et les prolongements (économiques, idéologiques, politiques, philosophiques, etc.)?
- e) Quels étaient les groupes d'intérêts en présence?

- f) Quelles stratégies suivaient ces groupes dans leurs interactions?
- g) Où peut-on situer les limites du système à étudier?
- h) Peut-on établir un modèle décrivant ce système et les tensions qu'il comportait?
- i) Quels sont les aspects, quelles sont les variables d'importance centrale dans ce système, du point de vue des problèmes posés?
- j) Quelles sont les personnes, les sources d'information, à consulter pour connaître l'évolution de ces aspects ou de ces variables centrales?
- k) Quelle a été l'évolution du système au cours de l'expérience, du point de vue des problèmes posés?
- l) Peut-on établir un modèle décrivant cette évolution?
- m) Peut-on expliquer cette évolution, en distinguant certaines nécessités?
- n) Quelle était l'attitude des personnes qui ont fourni les informations?
- o) Quels biais éventuels ont affecté les résultats recueillis?
- p) Quels critères d'évaluation paraissent pertinents pour chaque groupe?
- q) Quels ont été les résultats pour chacun de ces critères?
- r) Y a-t-il divergence de vues sur les faits eux-mêmes?
- s) Y a-t-il divergence de vues sur leur interprétation?
- t) Quelles sont les conclusions des divers groupes concernés par la réforme?
- u) Que proposent-ils pour la suite de l'expérience?
- v) Quels problèmes risquent de se poser en cas d'arrêt, de continuation, de généralisation de l'expérience?

5. Les rapports

Cette dernière étape de la recherche joue un rôle capital, du fait qu'elle conditionne en grande partie la compréhension et l'utilisation effective des informations recueillies et analysées précédemment. Son efficacité n'est pas seulement fonction de son style, cependant. Elle pose plutôt un problème d'adaptation à la demande. Le même rapport ne pourra pas répondre aux désirs des enseignants, des parents, des autorités ou des chercheurs. Le contenu (c'est-à-dire les types de problèmes examinés), comme la forme (et en particulier les termes techniques utilisés), doivent être choisis en fonction de l'attente de chaque public, faute de quoi le message ne passera pas.

L'explication détaillée des analyses statistiques effectuées, par exemple, est indispensable pour qu'un chercheur puisse apprécier la signification des résultats quantitatifs présentés. Pour le non-spécialiste, au contraire, les mêmes informations obscurcissent les conclusions de l'étude. Un administrateur, en particulier, ne pourra souvent rien tirer d'utile pour lui d'un rapport scientifique qui ne pose pas le problème en termes de décisions à prendre.

Trouver le mode de communication le plus approprié implique d'utiliser parfois d'autres supports d'information. Ecouter un compte rendu oral, suivre des conférences, participer à un colloque, reste pour beaucoup de gens la meilleure façon de prendre connaissance d'un domaine nouveau. Voir un film, ou pouvoir au moins s'appuyer sur l'image, est utile à d'autres. Il ne faut pas hésiter à consacrer des sommes peut-être considérables à mettre au point divers types de comptes rendus: le bon emploi de toutes les sommes dépensées précédemment pour la recherche en dépend souvent.

L'adaptation des rapports de recherche aux publics visés ne signifie nullement que les conclusions doivent varier d'un groupe à l'autre, ou que le rapport doit être écrit en vue de plaire à chaque interlocuteur. L'étude perdrait alors tout caractère scientifique et toute utilité. Heureusement, le simple fait que les divers publics concernés auront fatalement accès, tôt ou tard, aux divers rapports, même ceux qui n'étaient pas prévus pour eux, offre une garantie d'objectivité essentielle. Rien de ce qui aura été écrit n'a de raison de rester confidentiel, puisque le rapport vise seulement à refléter une réalité qui est accessible à tous. Vouloir cacher une partie de cette réalité à un groupe ou à un autre relè-

verait manifestement de la manipulation et de la mauvaise foi.

Ce principe de transparence, qui doit permettre à chacun de refaire la démarche suivie par l'évaluateur, ne suffit pas à résoudre toutes les questions d'éthique. Stake a présenté à Liège un document où plusieurs évaluateurs américains exposaient leur point de vue sur une série de problèmes délicats. Que faire si l'évaluateur découvre des faits qui mettent en cause l'honnêteté de telle ou telle personne? Que faire si des résultats publiés sont repris par d'autres d'une façon partielle, dans un contexte qui en fausse la signification? A qui appartient la recherche? - au chercheur? - aux maîtres, aux parents, aux élèves? - à ceux qui sont concernés et qui ont fourni les informations? - à l'autorité qui a commandé l'étude? - au public, qui a droit à la vérité dans une société démocratique et qui doit pouvoir contrôler la gestion de l'école? Il ne semble pas y avoir de réponse simple à ces questions. Une négociation entre les parties concernées est toujours nécessaire. Les principes directeurs dans cette négociation sont à rechercher à un niveau de généralité supérieur: celui des conditions de réussite d'une innovation pédagogique, problème largement étudié déjà dans les publications du CERI. C'est par une participation très large de tous à la gestion de l'innovation que celle-ci peut réussir, et il en est de même de l'évaluation pédagogique qui n'en constitue qu'un volet.

VIII. L'utilisation de l'évaluation

La discussion précédente donne le ton de ce que l'on peut dire sur l'utilisation à prévoir des conclusions du chercheur.

1. Evaluation opérationnelle

Dans la conception traditionnelle, la tâche du scientifique était terminée quand il avait déposé un rapport sur les faits observés. Bien souvent il s'agissait d'un texte de style scientifique, accessible uniquement aux spécialistes. Soit par manque de temps, soit par manque d'intérêt, soit enfin par naïveté et difficulté à concevoir les besoins de ses partenaires, le chercheur quittait le domaine lorsque ses analyses étaient terminées. Il fallait insister auprès de lui, par exemple, pour qu'il donne quelque information en retour aux enseignants et aux élèves qui avaient accepté de participer à sa recherche. Sa motivation étant la connaissance, il avait hâte d'aborder de nouveaux problèmes. L'exploitation pratique de ses résultats le préoccupait peu personnellement.

Si l'on examine la signification d'ensemble de l'entreprise d'évaluation, on voit pourtant qu'elle se justifie davantage par les adaptations concrètes qu'elle doit permettre d'apporter au système scolaire que par les conclusions théoriques généralisables qu'elle établit. En effet, ses conclusions sont trop souvent limitées par les caractéristiques particulières de la situation étudiée. Autrement dit, l'évaluation pédagogique est plus proche généralement de la recherche opérationnelle que de la recherche fondamentale. Il faut en tirer les conséquences et faire en sorte que le rapport final ne reste pas dans un tiroir ou sur un rayon de bibliothèque.

2. Evaluation participative

C'est bien l'ambition du nouveau style d'évaluation présenté ci-dessus que de s'intégrer à la pratique. Wulf (1975) parle d'une évaluation «proche de la pratique» qui doit permettre de gérer des innovations pédagogiques décentralisées. Il la considère comme un élément intégré dans le processus d'ensemble de développement de nouveaux curriculums. Comme telle, l'évaluation exige la coopération étroite des enseignants, des chercheurs et des autorités. Si tous doivent participer à la gestion de la réforme, ils doivent également participer à son appréciation. Ce n'est donc pas au chercheur à formuler un jugement. Son rôle est plutôt de recueillir les éléments permettant aux uns et aux autres de se faire une idée exacte de ce qui se passe, et une opinion sur les mesures à prendre en conséquence.

3. Evaluation formative

Ce retour d'information vise d'abord à permettre aux divers responsables d'adapter leur action au développement de la situation. Il n'est pas tellement nécessaire pour eux, qui font partie du groupe de lancement de la réforme, de se poser la question de sa continuation ou de son arrêt. En fait, on ne revient jamais en arrière. Il leur faut plutôt établir *comment* poursuivre au mieux l'évolution commencée. Les incertitudes et les conflits éventuels portent pour eux sur la tactique à suivre à l'avenir. Ils n'en sont pas encore à rechercher le jugement de l'histoire.

4. Evaluation à court terme

L'évaluation bilan traditionnelle laisse se dérouler tout le projet jusqu'à son achèvement pour pouvoir ensuite porter un jugement d'ensemble. Il semble en effet difficile de conclure avant que l'expérience ait pu développer tous ses effets. C'est ainsi que de multiples recherches américaines ont pris la forme d'études longitudinales, où l'on essayait de voir ce que devenaient les élèves à la suite du programme. Dans ce cas, les données à analyser s'accumulent pendant des années. Il n'est pas rare que le programme soit terminé, les responsables dispersés, quand le rapport vient à être publié, dans l'indifférence quasi-générale... L'évaluation formative, au contraire, implique de suivre l'expérience au jour le jour et d'en diffuser les résultats de façon continue. On a proposé à Liège l'établissement d'un poste de contrôle que centraliserait tout ce que l'on sait sur la réforme en cours. Cette formule permettrait de raccourcir les boucles d'adaptation en minimisant le délai causé par l'évaluation elle-même.

Conclusions

On peut, en terminant, prendre un peu de recul et chercher à évaluer l'évaluation élargie elle-même. Du point de vue pratique, il ne fait pas de doute que cette nouvelle perspective rendra beaucoup de services. Elle semble directement issue d'une réflexion sur les besoins concrets des responsables et d'un dialogue avec eux. Elle s'emploie efficacement à les satisfaire. Elle découle d'une politique générale de participation de tous les milieux concernés au pilotage des innovations pédagogiques. Elle est ainsi un instrument réfléchi et cohérent de gestion des systèmes scolaires en évolution.

Du point de vue théorique, cette conception nouvelle pose beaucoup de problèmes, mais c'est peut-être son grand mérite. La pédagogie expérimentale semblait en effet trop fixée, depuis longtemps, sur un modèle à la fois rigoureux et irréaliste. Cette contestation va rouvrir un débat de fond, trop longtemps esquivé.

Le premier problème, en effet, que pose cette nouvelle perspective est celui de l'objectivité. Ses détracteurs n'ont pas de peine à souligner le danger que représente l'appel à l'opinion subjective des intéressés pour fonder un jugement pédagogique. L'expérience d'Hawthorne, comme les travaux de Festinger sur la modification des attitudes, montrent que le simple fait d'être participant dans une expérience modifie la perception et l'attitude de celui qui est interrogé. Il faudrait donc, plutôt que de s'en remettre aux témoins, détromper, par la mise en évidence de faits objectifs, ceux qui s'illusionnent.

Ceux qui contestent le modèle scientifique traditionnel répondent que l'objectivité dont on le pare est, en fait, également illusoire. D'une part, comme on l'a vu, il ne permet pas vraiment de déterminer des lois générales: soit on ne peut pas contrôler les variables essentielles, soit on les contrôle si bien qu'on les oublie (en laboratoire) et les résultats n'ont plus alors de portée pratique. Plus profondément encore, on peut souligner le caractère «construit» et artificiel des théories scientifiques, surtout de toutes celles qui se fondent sur un modèle mathématique ou statistique élaboré. Il devient impossible de les vulgariser sans les trahir. On ne sait plus bien, par exemple, quel rapport les modèles d'ap-

prentissage, l'analyse factorielle des aptitudes, ou les statistiques multivariées entretiennent encore avec les problèmes concrets de la classe. N'est-il pas temps de revenir au vécu, de retrouver ainsi l'intégralité du réel, au lieu d'élever un échafaudage ambitieux qui finit par nous le cacher? Il semble que l'intuition directe puisse nous permettre un autre ordre de connaissance, infiniment plus riche que celui de la construction scientifique.

Mentionner l'intuition, c'est relancer un débat philosophique bientôt centenaire sur la possibilité ou la signification d'une science de l'homme. Par le simple fait qu'il peut prendre conscience des déterminismes qui l'enserrent, l'homme n'échappe-t-il pas toujours plus ou moins à la loi qui prétend le régir? Il semblerait qu'on en fasse l'expérience aujourd'hui dans les sciences de l'éducation.

L'évolution de la réflexion en pédagogie suit en effet un tracé analogue à celui que l'on observe actuellement parmi les autres sciences du comportement. Après un point de départ positiviste, la psychologie aussi est en crise. On voit aujourd'hui des psychopédagogues aussi fameux que Cronbach douter de la possibilité d'objectiver véritablement leur domaine (1974). Le behaviorisme est battu en brèche aux USA. Le modèle piagétien s'y impose. Il n'est donc pas étonnant qu'en pédagogie également on recherche de nouvelles voies. Le cadre conceptuel de la «compréhension», aidé des cadres structuraliste et fonctionnaliste, ne peut-il rendre mieux compte de la réalité humaine que le cadre positiviste?

Il semble difficile de répondre de façon assurée à cette question. En tout cas, d'autres sciences humaines commencent à apporter leur contribution à la pédagogie. Elles éclairent une série d'aspects de la réalité, partiels sans doute, mais complémentaires. La face interne des phénomènes retrouve place, enfin, dans un univers scientifique qui semblait l'avoir bannie pour toujours. Ceux pour qui les points de vue subjectif et objectif sont irrémédiablement étrangers espèrent ainsi prendre une revanche contre l'objectivité desséchante de la psychologie «en troisième personne». Pour d'autres, on peut aussi déceler dans cette évolution la recherche d'une synthèse, bien lointaine encore, où les lois objectives porteraient sur des variables dont la signification nous serait accessible par l'intuition. Ce ne sont là, aujourd'hui, que des conjectures. Il est préférable de ne pas engager l'avenir et de constater simplement un élargissement, désormais bien marqué, de nos conceptions scientifiques. La pédagogie expérimentale se dégage d'une norme trop rigide, qui faisait que toute information, non confirmée par un test statistique, n'était pas estimée digne de foi. D'autres sciences existent qui ont, elles aussi, leurs exigences d'universalité, et qui semblent pourtant plus souples, mieux adaptées aux problèmes concrets de l'évaluation pédagogique que les plans expérimentaux de Fisher. Sans renoncer aux acquis du passé, il faut donc nous ouvrir à cet enrichissement de nos moyens d'action, qui ouvre des perspectives passionnantes sur les plans de la recherche théorique comme de l'application pratique.

Die Erweiterung der Evaluation

Die pädagogische Forschung sollte zur ständigen Anpassung des Schulsystems beitragen; Voraussetzung dafür ist, dass sie in erster Linie den verantwortlichen Personen Lösungen auf ihre Fragen liefert, indem sie ihnen Informationen zur Verfügung stellt, die zur Begründung ihrer Entscheidungen nötig sind. Dies bringt eine Verschiebung der Akzente im Vergleich zur üblichen wissenschaftlichen Forschung. Nicht der Forscher evaluiert, sondern die am Versuch beteiligten Personen. Schliesslich beruht die neue Art der Evaluation nicht mehr auf dem Begriff einer positivistischen Wissenschaft, sondern stützt sich auf das Verstehen der Phänomene, die in ihrem subjektiven Referenzrahmen ins Auge gefasst werden. Der Versuch wird zur Aktionsforschung, in der der Forscher selbst ein Element der von ihm beobachteten Situation ist. Objektivität bedeutet intersubjektive Übereinstimmung der an der Reform teilhabenden Personen.

The Expansion of Evaluation

Educational research should contribute to the continual adjustment of the school system. This means that it must provide the information asked for by those in charge enabling them to take the necessary decisions. This means a shift of emphasis compared to the usual experimental one. It is no longer the researcher who evaluates, or who decides which factors should be considered, or in relation to which values an appreciation must be made. This must be done by the people involved in the experiment, i.e. the people who take the decisions. This new approach to evaluation is based on a conception of science no longer positivist but on the understanding of phenomena considered in their subjective framework. The experiment is transformed into an «action-research» where the researcher becomes one of the elements he is observing. Objectivity becomes an intersubjective agreement between the partners of the reform.

BIBLIOGRAPHIE

- Argyris, Ch.: Intervention theory and method: a behavioral science view. Reading (Mass), 1970.
- Brunner, J. S. et Olson, D. R.: Apprentissage par expérience directe et apprentissage par expérience médiatisée. In: Perspectives, Vol. 3, No. 1, 1973, p. 21 - 42.
- De Bruyne, P., Herman, J., de Schoutheete, M.: Dynamique de la recherche en sciences sociales. Paris, 1975, p. 202 - 206 (Coll. Puf, section Le Sociologue).
- Campbell, D. T.: Qualitative Knowing in Action Research. Exposé présenté à l'A.P.A., New Orleans, 1974. Occasional papers of the Stanford Evaluation Consortium, Stanford Univ.; to appear in The Journal of Social Issues.
- Cronbach, L. J.: Beyond the two disciplines of scientific psychology. Exposé présenté à l'A.P.A., New Orleans, 1974, Occasional papers of the Stanford Evaluation Consortium, Stanford University.
- Cronbach, L. J.: Evaluation for course improvement. In: Heath, R. W., (Ed), New curricula, New York, 1964.
- Glaser, R.: Intelligence, learning and the new aptitudes. In: John Leedhaw (Ed.), Aspects of Educational Technology, No. VII, Bath (U.K.), 1973, p. 29 - 45.
- Haag, D.: Gestion d'un système scolaire - Perspectives globales. Thèse de la Faculté de Droit et de Sciences Economiques, Politiques et Sociales de l'Université de Neuchâtel, Neuchâtel, 1975.
- Huberman, M.: Evaluating the effectiveness of schooling. In: International Review of Education, 3/1973, p. 356 - 371.
- Munari, S.: Gestion d'un système scolaire - Fonctionnement interne du système. Thèse de la Faculté de Droit et de Sciences Economiques, Politiques et Sociales de l'Université de Neuchâtel, Neuchâtel, 1975.
- Niedermann, A. et Santini, B.: Die Evaluation eines neuen Schultyps am Beispiel der Weiterbildungsschule Zug (Schweiz). In Frey, K., in Zusammenarbeit mit Achtenhagen, F. (Hrsg.): Curriculum-Handbuch. 3 Bände, München, 1975.
- Parlett, M.: The new evaluation. In: Trends in Education, No. 34, 1974, p. 13 - 18.
- Scriven, M.: The methodology of evaluation. In: R. Stake (Ed.), Perspectives of curriculum evaluation. Chicago, 1967.
- Smith, L. et Pohland, D. P.: Education, Technology and the Rural Highlands. In: Four Evaluation Examples: anthropological, economic, narrative, and portrayal, Chicago, 1974, p. 5 - 54 (AERA Monograph No. 7).

Weiss, R. S. et Rein, M.: The evaluation of broad - aim programs: experimental design, its difficulties and an alternative. In: Admin. Science Quart., 1970, 15 (1), p. 97 - 112.

Wulf, Ch.: Evaluation im Rahmen praxisnaher Curriculumentwicklung. In: Isenegger, U. und Santini, B.: Begriff und Funktionen des Curriculums. Weinheim und Basel, 1975, p. 131 - 159. (Beltz Monographien Erziehungswissenschaft. Studien und Forschungsberichte aus dem Pädagogischen Institut der Universität Freiburg (Schweiz), 12).