

CHOISIR LA DEMARCHE D'EVALUATION QUI CONVIENT

Jean CARDINET

RECHERCHES
89.109 - Octobre 1989

CHOISIR LA DEMARCHE D'EVALUATION QUI CONVIENT

Jean CARDINET

IRDP

INSTITUT ROMAND DE RECHERCHES ET DE DOCUMENTATION PÉDAGOGIQUES
Faubourg de l'Hôpital 43 CH 2000 Neuchâtel Tél. (038) 24 41 91

CARDINET, Jean. - Choisir la démarche d'évaluation qui convient / Jean Cardinet. - Neuchâtel : Institut romand de Recherches et de Documentation Pédagogiques, 1989. - 12 p. ; 30 cm. - (RECHERCHES ; 89.109)
Bibl. p. 12.

Evaluation	Suisse romande
Epistémologie	
Méthodologie	
Programme de formation	
Formation d'adultes	
Méthode pédagogique	

La reproduction, totale ou partielle, des publications de l'IRDP est en principe autorisée, à condition que leur(s) auteur(s) en ai(en)t été informé(s) au préalable et que les références soient mentionnées.

CHOISIR LA DEMARCHE D'EVALUATION QUI CONVIENT

Jean CARDINET

Résumé

Trouver la meilleure méthodologie d'évaluation n'est possible que si l'on se donne le cadre de référence pour en juger. Or, ce cadre dépend de la réponse donnée à trois questions.

- (1) Sur quelle base accepte-t-on une affirmation comme vraie : généralité du phénomène, cohérence de son explication, ou accord intersubjectif des observateurs ?*
- (2) Quelle sorte de décision l'évaluation doit-elle fonder : régulation proactive, interactive ou rétroactive ?*
- (3) Privilège-t-on la sûreté de la connaissance du cas particulier étudié, ou la validité des prédictions qu'on peut en tirer ?*

Ces questions sont si fondamentales qu'elles déterminent la méthode à suivre pour évaluer aussi bien des personnes, que des interventions, ou des institutions.

WELCHE MASSNAHMEN FÜR EINE ANGEMESSENE BEWERTUNG ?

Jean CARDINET

Zusammenfassung

Will man die geeignetste Methode zur Bewertung finden, muss zuerst einmal festgelegt werden, worauf sich die Beurteilung beziehen soll. Die Rahmensetzung hängt von der Antwort auf die drei folgenden Fragen ab.

- (1) Welches ist die Basis, auf der eine Aussage als richtig angesehen wird: Allgemeinheit des Phänomens, Kohärenz in der Erklärung, intersubjektive Übereinstimmung der Beobachter ?*
- (2) Welche Art des Entscheids soll die Bewertung begründen: proaktive, interaktive oder retroaktive Regulation ?*
- (3) Worauf wird mehr Wert gelegt: auf sichere Kenntnisse in einem gelernten speziellen Gebiet oder auf die Gültigkeit der daraus zu ziehenden Voraussagen ?*

Diese Fragen sind so grundlegend, dass sie die Wahl der Methode bestimmen, nach der Personen, Interventionen oder Institutionen bewertet werden.

CHOISIR LA DEMARCHE D'EVALUATION QUI CONVIENT *

Jean Cardinet

1. NECESSITE D'UNE STRATEGIE EXPLICITE

Le titre initialement proposé pour cette contribution : "Evaluer un programme de formation ... constitutionnellement", voulait mettre une note d'humour dans un sujet austère, en restant dans le style "révolutionnaire" choisi pour le Congrès de l'ADMEE de 1989. L'essentiel du titre était donné par le mot "constitutionnellement", qui évoquait l'idée d'un contrat social, canalisant les conflits de pouvoir dans un cadre de procédures convenues. Il s'agissait en fait de préciser selon quelles démarches il fallait conduire un projet d'évaluation, mais en cherchant surtout à éviter que les évaluateurs ne prennent leurs options de façon purement arbitraire. L'important était de trouver des règles pour fonder leurs choix stratégiques initiaux, étant entendu que leurs décisions ultérieures, concernant notamment la création et la mise au point d'instruments, en découleraient très naturellement et de manière plus contrôlable.

Les ouvrages actuels sur la méthodologie de l'évaluation (par exemple, Nadeau, 1981) se contentent de présenter l'éventail des approches possibles, en indiquant bien leurs avantages et leurs inconvénients respectifs, mais de manière très générale, sans préciser dans quelles conditions telle méthode peut convenir mieux que telle autre. Cronbach lui-même, dans ses deux importants ouvrages méthodologiques (1980, 1982) recommandait bien de combiner plusieurs approches, mais laissait à l'intuition de l'évaluateur le soin de trouver le dosage approprié, en fonction de chaque situation.

Il est pourtant assez évident que toutes les décisions tactiques, concernant notamment la meilleure façon de recueillir des données, sont d'intérêt secondaire en regard des décisions stratégiques, relatives par exemple à la confiance que l'on décide d'accorder au sentiment de satisfaction des personnes qui ont participé au projet évalué (veut-on, ou non, s'appuyer sur leurs perceptions subjectives ?). De même, il est essentiel de savoir si l'évaluation porte sur le bilan final de l'opération évaluée, ou sur la façon dont elle a été conduite, ou enfin sur les risques de l'entreprise, liés à son contexte initial. Toute la démarche d'évaluation dépendra, bien évidemment, de la réponse apportée à ces questions.

C'est pourquoi la discussion portera ci-dessous sur trois options stratégiques qui paraissent si fondamentales qu'elles semblent même pouvoir être prises indépendamment de l'objet à étudier, qu'il s'agisse d'individus, de programmes de formation, d'institutions ou d'innovations. Les exemples abordés se centreront cependant sur des cas d'évaluation d'actions de formation.

* Contribution présentée au Colloque de l'ADMEE 1989, à Châtenay-Malabry, le 27 septembre 1989.

2. CHOIX D'UNE EPISTEMOLOGIE

Le premier embranchement du point de vue méthodologique porte sur le cadre de pensée, le "paradigme" selon De Bruyne, Herman et De Schoutheete (1974), c'est-à-dire la façon dont on admet qu'on peut connaître la réalité. Il semble qu'aujourd'hui trois approches épistémologiques majeures sont à considérer: objective, subjective et interactive.

2.1. Approche objective

La première correspond à la conception de la démarche scientifique qui est admise et mise en pratique dans les sciences de la nature : physique, biologie, etc. Elle vise à détecter des relations régulières entre les phénomènes, à formuler des lois, en isolant l'effet de certaines variables indépendantes par rapport à d'autres variables dépendantes.

L'expérience scientifique conduit alors à des résultats probants, si sa validité interne est indiscutable, c'est-à-dire si les observations recueillies sur un échantillon nécessairement limité permettent néanmoins une inférence assurée à tout le domaine de validité choisi au départ.

(Quant à savoir si dans une autre circonstance ultérieure, différant du domaine étudié expérimentalement, le même résultat serait aussi à attendre, il s'agit là d'une autre question, qui relève de la validité externe de l'expérience. Les deux validités sont en partie contradictoires et l'évaluateur doit faire un choix de priorité entre les deux, comme on le verra plus bas).

Le facteur à relever ici, c'est que la subjectivité de l'évaluateur constitue la première source de variation que ces sciences tiennent absolument à supprimer, parce qu'elle risque de menacer la validité interne de l'étude : il serait trop facile de mettre en doute les conclusions, si elles reposaient tant soit peu sur le jugement personnel d'un observateur dont la neutralité ne pourrait être assurée.

L'accord des experts ne peut être obtenu, dans cette conception, que si les conclusions découlent de relevés objectifs, permettant de contrôler sans doute possible que les critères fixés au départ ont été atteints. Ainsi, l'efficacité d'un programme sera prouvée si l'on parvient à opérationnaliser ses objectifs pédagogiques et à s'assurer que les apprenants dépassent bien, à l'examen final, le seuil de suffisance requis.

2.2. Approche subjective

La deuxième approche se présente comme l'antithèse de la première. Elle souligne l'artificialité du découpage effectué dans la réalité, qui rend illusoire l'espoir de pouvoir généraliser de la situation expérimentale aux situations d'application ultérieures; elle dénonce donc le manque de validité externe de l'approche précédente.

Plus fondamentalement, elle affirme que la raison de ce manque de pertinence des travaux expérimentaux provient justement de ce qu'ils ne prennent pas en compte les facteurs subjectifs. Bien loin de les éliminer, il faudrait les considérer en priorité, car les hommes réagissent moins aux faits qu'aux significations qu'ils accordent à ces faits.

Si des spécialistes d'orientations diverses peuvent néanmoins s'accorder sur certains constats historiques, cela démontre que la valeur de vérité peut être aussi élevée dans ce paradigme subjectif que dans le précédent : seules les démarches pour l'établissement de la vérité doivent être transposées, comme l'a montré Guba (1981). Après tout, dans le domaine judiciaire, un degré suffisant de validité interne peut bien être obtenu pour les conclusions d'un détective, sans recourir à des tests statistiques.

Mais cette approche oblige à prendre en compte le rôle propre de l'évaluateur, qui ne peut s'extraire de la situation et qui influence les répondants, par le fait même qu'il s'adresse à eux. A la limite, la notion de vérité comme accord entre la pensée et la réalité tend à disparaître, car il n'existe plus de terme fixe que l'on puisse appeler le réel. Seules existent des représentations diverses en interaction, et donc en évolution continue. Une régulation de la situation semble plus facilement envisageable qu'un bilan dans ces conditions (comme on le verra aussi plus loin), car il y aurait autant de bilans à faire que de sujets.

2.3. Approche interactive

La troisième approche s'efforce de dépasser les limitations des deux précédentes en acceptant un cadre de pensée plus global, où la réalité est à la fois objective et subjective. Sans doute, ce qui est donné d'abord à l'observateur, c'est la réalité subjective de points de vue souvent contradictoires. Cependant, la confrontation de ces perceptions différentes met en marche une suite de restructurations des représentations des uns et des autres, qui tendent alors vers une vision plus large, et partant plus stable, du réel.

Les instruments d'observation et les critères d'évaluation sont choisis à partir des préoccupations diverses des personnes concernées par le programme. Dans la mesure où un aspect de la situation est considéré comme important par l'un des partenaires, il doit être pris en compte dans la démarche d'évaluation des autres. Mais il n'est généralement pas nécessaire d'unifier les poids accordés à chaque critère par les différents acteurs. C'est déjà un progrès vers l'objectivité chez l'évaluateur de pouvoir décrire et faire comprendre les réactions particulières de chaque groupe. Un bilan qui occulterait leurs divergences ne permettrait pas aux responsables d'élaborer une politique pour l'avenir en connaissance de cause.

2.4. Sur quoi fonder le choix ?

Ce n'est pas le caractère plus englobant de la troisième approche qui peut suffire à justifier qu'on la préfère aux autres. Elle ne formule pas nécessairement d'évaluation plus valide et elle ne conduit pas nécessairement à des décisions plus efficaces. Il suffit de comparer la chirurgie, la psychanalyse et la médecine psychosomatique qui correspondent aux trois mêmes épistémologies, pour comprendre que les trois approches ont chacune leur champ d'application spécifique. Parler d'une supériorité générale de l'une ou de l'autre n'aurait aucun sens. Par contre, il est évident que, selon la situation, en évaluation comme en médecine, c'est le recours aux méthodes des sciences naturelles, humaines, ou psychosociales, qui peut se révéler le plus approprié. Seule l'analyse de la demande et de son contexte peut permettre de fonder le choix entre les trois approches du réel.

Il semble qu'une évaluation externe soit requise dans une situation de conflit où les parties s'en remettent, pour les départager, au jugement des faits objectifs rassemblés et analysés par un expert impartial. Ce qui est attendu dans ce cas, c'est la formulation d'une connaissance assurée ou d'une règle systématique, en tout cas une réponse générale à un problème général: telle formation permet d'atteindre tel niveau de compétence, en tant de temps, par exemple. L'application de cette "loi scientifique" au cas particulier qui a suscité le besoin d'évaluation relève d'une autre logique, et souvent les décideurs se réservent d'exploiter comme bon leur semble, voire de rejeter comme non appropriée à leur cas, la conclusion "objective" de l'évaluateur externe.

L'évaluation interne est celle qui relève, justement, de la logique des responsables de l'innovation, ou du programme de formation en question. Que telle façon de faire conduise en général à tel résultat ne prouve pas que, dans le cas particulier dont ces responsables ont la charge, la liaison de cause à effet fonctionne toujours: les personnes en cause, le contexte socio-politique, etc., peuvent modifier cette relation. Les décideurs doivent donc prendre en compte toute la réalité, lorsqu'ils ont à intervenir sur le plan social. Evaluer consiste pour eux à mettre en lumière tous les aspects de la décision à prendre. Ceci implique de prévoir les interventions des autres acteurs sociaux. Comment mieux y parvenir qu'en les interrogeant pour comprendre leurs perceptions et leurs intentions ? L'approche subjective paraît alors entièrement justifiée dans cette évaluation centrée sur la décision.

L'évaluation négociée se présente finalement comme un élargissement des deux précédentes, extension qui est sans doute souhaitable dans une société où la concertation des décisions devient une règle de fonctionnement de plus en plus explicite. Cependant la négociation d'un compromis comme base de la décision n'est pas strictement indispensable (lorsque la faiblesse des enjeux, ou l'urgence des décisions à prendre, ne justifient pas qu'on y consacre du temps, par exemple).

Dans l'idéal, ces trois phases pourraient s'appuyer l'une sur l'autre. La recherche "objective" de faits pourrait servir de point de départ à une deuxième phase, celle des interprétations multiples, liées aux positions différentes des partenaires interrogés. La troisième phase serait celle où ces visions subjectives seraient confrontées, pour construire une conception d'ensemble de la réalité, capable d'intégrer les divergences de points de vue dans la préparation de décisions répondant de façon synthétique aux demandes des uns et des autres.

D'un point de vue strictement logique, on objectera peut-être qu'il est possible de dissocier la position de l'évaluateur (externe, interne, ou intermédiaire) du cadre de référence épistémologique qu'il utilise. Dans les faits, la liaison suggérée plus haut se révèle pourtant à la fois fréquente et naturelle.

3. CHOIX D'UNE FONCTION

3.1. Les trois perspectives temporelles

Le deuxième embranchement fondamental est celui que Scriven avait explicité dès 1967, en montrant la différence radicale de logique entre une évaluation-bilan, tournée vers le passé, et une évaluation-régulation, visant les décisions adaptatives à prendre dans le présent. Cette amorce de typologie suggérait naturellement (Cardinet, 1977, 1986) l'existence d'une finalité pronostique: en effet, l'anticipation des situations futures permet de fonder la régulation "proactive" (selon l'expression de Allal, 1988), indispensable au fonctionnement de tout système. Elle correspond à une évaluation tournée vers l'avenir et servant à s'orienter par rapport aux divers possibles.

Cette distinction entre les trois fonctions d'orientation, de régulation et de certification a permis de distinguer trois formes d'évaluation: prédictive, formative et sommative. Ces concepts sont trop connus aujourd'hui pour qu'on ait besoin de les définir à nouveau. Mais il faut peut-être rappeler que les trois types de finalité, correspondant à trois types de boucles cybernétiques, doivent s'appuyer sur trois ordres de faits différents. Les options prises influencent donc le type d'observations à effectuer.

Pour un bilan, c'est la situation à un moment donné qui est à prendre en compte, par une sorte de prise de vue en instantané, qui fige le mouvement, de manière à permettre de totaliser les résultats sous forme d'actif et de passif. Les tendances, les démarches en cours sont ignorées, seuls les produits réalisés sont considérés. L'observation portera donc en priorité sur tout ce qui est acquis, par les individus ou les institutions, et d'abord sur ce qui peut être compté ou certifié: diplômes, réussites et échecs, position financière, etc. Seule la comparaison de bilans successifs peut rendre un aspect dynamique à ces constats qui prennent la forme, selon l'épistémologie choisie, de résultats d'expériences, de jugements historiques, ou de rapports d'audits participatifs.

Pour une gestion en continu, au contraire, les points de repère sont à prendre dans l'étude des processus en cours, qu'il faut surveiller parce que ce sont eux qui déterminent les résultats. Le groupe social constituant l'institution fonctionne-t-il convenablement ? L'enseignement se déroule-t-il de façon satisfaisante ? Les apprenants suivent-ils les stratégies d'apprentissage appropriées ? Les observations vont devoir devenir plus qualitatives que pour un bilan de résultats et les conclusions, pour être probantes, devront être davantage étayées de contre-expertises diverses. Dans l'optique d'une recherche tournée vers la mise en évidence de relations objectives, c'est la conformité du fonctionnement observé au déroulement prévu que l'on cherchera à contrôler. Dans une optique subjectiviste, on suivra

l'évolution des opinions des acteurs. Si l'approche, enfin, est interactive, l'observation de la pratique donnera les éléments objectifs sur lesquels les points de vue des acteurs pourront ensuite s'affronter.

Pour une évaluation prédictive qui doit fonder le choix entre plusieurs programmes d'action, ce sont les potentialités ou facteurs déterminants qu'il s'agit d'estimer, les appuis sur lesquels on peut compter et les obstacles à prévoir, pour chacun des programmes envisagés. On s'attachera donc à observer les caractéristiques favorables et défavorables du contexte, les aptitudes et les intentions des participants, etc., de façon à permettre une expérience de simulation dans l'optique objectiviste, une analyse des motivations dans l'optique subjectiviste, ou une étude prospective négociée dans l'optique interactive.

3.2. Le choix de priorité

Il est évident que c'est l'analyse de la demande d'évaluation qui permet de clarifier quel doit être le souci primordial de l'évaluateur et de déterminer en conséquence s'il faut étudier de préférence les résultats, les processus, ou les déterminants de l'opération en cours, ou bien plusieurs de ces niveaux à la fois, selon une pondération à déterminer.

Une première caractéristique décisive de la demande, qu'il faut faire préciser, est le degré d'attention porté aux causes de la situation observée. Dans l'optique objectiviste, il est indispensable de les contrôler, pour qu'on puisse vraiment attribuer les résultats obtenus à l'influence du programme et non à d'autres facteurs accidentels. Par exemple, il est nécessaire de voir comment les enseignants appliquent la méthode nouvelle qu'on leur a demandé d'utiliser, faute de quoi on ne pourrait savoir si des résultats insuffisants sont dus à cette méthode, ou au contraire à son mauvais usage. Dans l'optique subjectiviste, ce contrôle est superflu, puisqu'il suffit d'interroger les enseignants à la fois sur leur acceptation de la méthode et sur ses résultats. Dans l'optique interactionniste, la recherche des causes peut n'être abordée que si un conflit d'opinion apparaît à leur propos.

Une deuxième condition décisive pour le choix d'une évaluation tournée vers le passé, le présent, ou l'avenir, est l'importance des moyens matériels à disposition. Si le temps laissé pour exécuter le travail est trop court, il est évidemment impossible de suivre la succession des événements, depuis les causes jusqu'à leurs conséquences. De telles études longitudinales sont aussi généralement coûteuses, par le nombre de prises d'informations qu'elles supposent avant, pendant, et après le programme. Les études-bilans, surtout selon la méthode expérimentale, sont donc des entreprises importantes et difficiles à mettre sur pied. Une évaluation régulatrice visant une gestion à plus court terme paraît beaucoup plus économique, surtout dans l'optique subjectiviste. Une étude prospective ne demande parfois qu'un travail de commission.

Bien que dans l'idéal la collecte du maximum d'informations soit toujours souhaitable, en pratique les ressources en temps, argent et compétences sont toujours limitées. On peut donc comprendre que des évaluations soient engagées, qui ne considèrent que les résultats apparents sans en contrôler les causes, ou qui ne s'appuient que sur des informations subjectives. L'augmentation de coût qu'entraîneraient des enquêtes où les biais seraient mieux contrôlés rend en effet souvent de telles études irréalisables. En l'absence de contestation réelle de

l'impartialité des points de vue exprimés, pourquoi les évaluateurs devraient-ils les mettre en doute ? Les responsables des innovations, comme les politiciens, sont souvent obligés de ne traiter les problèmes que lorsqu'ils apparaissent.

4. CHOIX D'UNE VALIDITE

4.1. Les deux étapes de l'évaluation

L'idée précédente, qu'on puisse accepter de conclure "sous réserve" un rapport d'évaluation, a quelque chose de scandaleux au premier abord: la tâche de l'évaluateur n'est-elle pas justement de lever de telles ambiguïtés ? Il faut bien reconnaître, pourtant, que les recherches d'évaluation sont presque toujours centrées sur une décision et que les études de ce genre ont encore moins la possibilité que les recherches centrées sur la connaissance de conclure de façon certaine.

Qui s'intéresserait en effet à évaluer un programme de formation, ou n'importe quelle autre activité sociale, s'il ne percevait pas à la clé la décision de poursuivre, modifier ou abandonner cette action ? Même si on accepte la position la plus objectiviste, dans la première des options présentées plus haut, le raisonnement implicite de l'évaluateur comporte donc deux étapes: d'abord estimer quelle est la situation véritable à partir des informations qu'il a obtenues, ensuite apprécier en quoi cette situation est pertinente par rapport à la décision à prendre. Ces deux questions correspondent respectivement à la validité interne et à la validité externe de sa démarche évaluative.

4.2. Fondements de la validité interne

S'assurer que la réalité correspond bien à ce que l'on en dit pose un problème d'inférence dans quatre domaines, si l'on reprend la catégorisation quadripartite proposée par Cronbach (1982): les unités étudiées, les traitements effectués, les observations recueillies, et les situations abordées. Ce n'est que si l'inférence est légitime dans ces quatre domaines qu'il est possible de considérer l'étude comme répétable de quelque façon, faute de quoi elle n'aurait aucune valeur de vérité démontrable et personne ne pourrait lui faire confiance.

Les unités étudiées doivent être un échantillon représentatif d'une population particulière, d'apprenants, d'enseignants, d'écoles, de régions, etc. La façon dont a été effectué l'échantillonnage est essentielle pour permettre une généralisation à la population correspondante. Si ce sont seulement des enseignants volontaires et convaincus d'avance qui ont mené l'expérience, comment généraliser leur degré d'acceptation à celui d'autres groupes de maîtres, par exemple ? Si les résultats ne sont connus que pour les élèves qui ont surmonté tous les obstacles au cours du traitement, comment parler de l'effet général de la formation ? Le choix des classes expérimentales est également un facteur critique. La liste serait longue de tous les biais à éviter dans ce domaine.

Les traitements que l'on cherche à évaluer sont aussi à considérer comme un échantillon d'une population de traitements qui doit être bien définie. Jamais

aucune expérience ne permettra d'appliquer la méthodologie idéale rêvée par l'initiateur d'une méthode; les applicateurs (et l'initiateur lui-même) déforment toujours le modèle. Jusqu'où ira-t-on pour considérer que la méthode a été appliquée, et non pas trahie ? Si le principe pédagogique est justement d'adapter la méthode aux apprenants et à la situation, quel type d'adaptation est licite ? Sans réponse à de telles questions, on ne sait pas ce qu'on doit observer. Comment parvenir alors à des conclusions répétables ?

L'échantillon des observations pose le problème de la généralisabilité d'une méthode de mesure à une autre, d'un évaluateur à l'autre, entre questions, entre formes de réponse, etc. L'absence de biais et la validité de "construct" des instruments est aussi à vérifier. La validité des observations pose également des problèmes qualitatifs, liés à la compréhension de la formulation (verbale, graphique) utilisée, à l'interprétation du sens de l'examen, à la perception du rôle de l'observateur, au traitement des données manquantes, etc.

Le choix des situations expérimentales est souvent difficile, mais pourtant l'univers de circonstances à propos desquelles on peut formuler une conclusion doit être soigneusement défini. A quelle étape d'apprentissage se réfère-t-on ? Le type de comportements que l'on peut observer chez un maître au cours d'une leçon d'épreuve, ou lors d'une visite de l'inspecteur annoncée de longue date, n'est pas nécessairement en rapport avec son comportement journalier. L'essentiel est de ne pas généraliser induement d'une étape à l'autre, ou d'une situation à l'autre.

Des livres entiers ont été consacrés à analyser les différentes sources d'erreur affectant l'interprétation de ce qu'on peut observer au cours d'une expérience (Wolins, 1982). Les exemples précédents ne visaient qu'à évoquer l'ensemble des problèmes d'interprétation qui affectent la validité interne d'une entreprise d'évaluation.

4.3. Fondements de la validité externe

Les inférences dont il vient d'être question sont d'autant plus justifiées que l'expérience est mieux contrôlée, c'est-à-dire que l'échantillonnage d'unités, de traitements, d'observations et de situations se fait dans des populations plus étroites, et par conséquent plus homogènes et mieux prédictibles.

Cette qualité se transforme en défaut dès que l'on passe à la seconde étape de la démarche d'évaluation, celle qui implique de porter un jugement valable pour la situation de décision envisagée. Jamais la réalité sur laquelle on doit intervenir n'est en effet identique à la réalité observée. Une extrapolation est toujours nécessaire, ne serait-ce que sur la dimension du temps, mais le plus souvent sur plusieurs des quatre domaines évoqués. Préciser point par point en quoi le contexte de la décision se distingue de celui de l'expérience permet néanmoins d'apprécier intuitivement le risque que l'on prend à effectuer une telle extrapolation.

Au moment où l'on planifie une expérience d'évaluation, la question se pose d'opter pour une définition étroite ou large de chacune des populations (des personnes, traitements, instruments, etc.). Le plus souvent, ce que l'on gagne du point de vue

validité interne provoque une perte du point de vue validité externe. Il vaudrait mieux, pour fonder une décision de généralisation, par exemple, avoir expérimenté le programme sur des groupes d'apprenants divers, avec des enseignants non convaincus au départ, etc. Le choix doit donc être fait de manière raisonnée dès le départ.

4.4. Bases du choix

Deux caractéristiques de la situation d'évaluation peuvent infléchir la définition des populations à échantillonner dans un sens restrictif, ou au contraire extensif.

Si le problème principal semble être la mise en cause du bilan de l'expérience en cours, l'effort doit porter, au cours de l'investigation, sur tout ce qui peut lever cette incertitude. Une augmentation de la précision des mesures, un contrôle plus serré des sources de variation imprévues, de façon à limiter les possibilités d'explications concurrentes, représentent une stratégie appropriée pour faire face à cette situation.

Si le problème prioritaire est au contraire de permettre une prédiction raisonnable de l'effet à attendre d'une série de décisions envisagées, en fondant mieux les extrapolations nécessaires, le devoir de l'évaluateur est d'élargir la base de données qu'il constitue, en faisant varier les conditions de sa prise d'information. En effet, si les observations qu'il recueille sont suffisamment diversifiées, il pourra ensuite observer les conséquences de ces décisions pour chaque sous-population séparément, ou du moins limiter l'ampleur des extrapolations à effectuer.

5. ANALYSE DE LA DEMANDE

Ces quelques points de repère pour choisir une démarche d'évaluation, qui convienne aux circonstances dans lesquelles une décision doit être prise, soulignent la nécessité d'une phase préalable d'investigation qualitative, visant à expliciter les besoins prioritaires. Comme la principale difficulté est alors de dépasser l'expression initiale de la demande, telle qu'un responsable peut la formuler en répondant spontanément à la question "Pourquoi ?", il faut prévoir une première phase exploratoire, plutôt divergente. Le choix des priorités correspondra ensuite à une phase d'orientation inverse, plutôt convergente, où le plan de l'étude devra être choisi de façon à recueillir le maximum d'informations au moindre coût.

5.1. Phase divergente

Par définition, on ne peut cadrer et programmer le questionnement initial. La méthode clinique suppose en effet qu'on prenne du recul par rapport à chaque information pour réorienter les prises d'information suivantes. Une liste de questions, ainsi que des réponses à leur trouver, peut cependant avoir une valeur heuristique pour l'évaluateur.

Questions

Quelle est la demande initiale ?
 Requête de qui ?
 Pour qui ?
 Qui est consulté ?
 En vue de quoi ?
 Quels enjeux ?
 Points d'aboutissement ?
 Objectifs intermédiaires
 A propos de quoi ?
 Place de l'évaluateur ?
 Légitimité de son intervention ?
 - du point de vue éthique et social
 - du point de vue légal et administratif
 Délimitation du domaine d'étude ?

 Formulation élaborée du problème ?

Réponses

Première formulation du problème
 Mandataires officiels et officieux
 Destinataires de l'information
 Partenaires explicites et implicites
 Scénarios possibles
 Intérêts de chaque groupe
 Décisions à prendre
 Stratégies des acteurs
 Sources de difficultés prévisibles
 Rôle qu'on veut lui faire jouer
 Explicitation
 - des règles du jeu contractuelles
 - du rattachement institutionnel
 Populations d'unités, de traitements, d'observateurs
 et prises en compte de situations
 Hypothèses de travail principales
 Aspects secondaires à contrôler
 Objections à prévoir
 Contrôles nécessaires

5.2. Phase convergente

Compte tenu de l'impossibilité de tout faire, il faut trouver le moyen de fixer des priorités, puis négocier un projet en tenant compte des contraintes de la réalité. La liste de questions ci-dessous peut aussi aider à la réflexion :

- quelles sont les questions les plus importantes pour fonder les décisions envisagées ?
- quelles sont celles pour lesquelles la réponse est la plus incertaine ?
- pour quelles hypothèses les contre-arguments semblent-ils les plus forts ?
- quelles informations sont les plus faciles à obtenir ?

Les questions satisfaisant les deux premiers critères devraient avoir la priorité. Mais les possibilités d'action, en termes de budget, de personnel et de temps disponible fixent un cadre limitatif à ces projets d'étude.

Le mode de publication et la propriété de l'information devraient aussi faire l'objet de contrats précis avant le début du travail.

6. OBJECTIF FINAL

Wilensky (1967, p. IX, cité par Cronbach, 1982, p. 11) a déjà défini les caractéristiques d'une bonne étude d'évaluation. Une telle recherche doit être :

"claire, c'est-à-dire compréhensible pour ceux qui doivent en tirer parti;

opportune, en fournissant ses conclusions au moment où elles peuvent être prises en compte;

fidèle : au sens où divers évaluateurs, en utilisant la même démarche, arrivent aux mêmes résultats;

valide, les concepts utilisés et les mesures obtenues ont prise sur la réalité, ce qui se vérifie par leur cohérence interne, leur valeur prédictive et leur accord avec les autres connaissances, obtenues indépendamment;

adéquate : prenant suffisamment en compte toute la réalité, en décrivant aussi le contexte de l'action, de l'événement, ou de la vie, des personnes ou groupes considérés;

de large envergure : par le fait qu'elle envisage l'ensemble des politiques possibles, ou qu'elle propose de nouveaux objectifs. "

En plus de ces critères relativement formels, on peut dire que le but de l'évaluation d'un programme est d'obtenir des informations qui permettent aux membres des groupes sociaux en présence de résoudre leurs conflits intelligemment. Ceci implique qu'ils reçoivent des indications les amenant à confronter leurs points de vue, à remettre en cause leurs préconceptions, et à construire ensemble des représentations plus convergentes de la réalité. Les informations fournies par l'étude d'évaluation doivent donc finalement guider leur nouvelle conceptualisation de la situation et réduire leur incertitude quant à la décision à prendre sur cette base.

* * *

Références

Allal, L. Vers un élargissement de la pédagogie de maîtrise : processus de régulation interactive, rétroactive et proactive. In: M. Huberman (Ed.) Assurer la réussite des apprentissages scolaires ? Les propositions de la pédagogie de maîtrise. Neuchâtel: Delachaux et Niestlé, 1988, p. 86-126.

Cardinet, J. Objectifs éducatifs et évaluation individualisée. Neuchâtel: Institut romand de Recherches et de Documentation Pédagogiques, 1977. IRDP/R 77.05.

Cardinet, J. Évaluation scolaire et mesure. Bruxelles: De Boeck, 1986.

Cronbach, L.J. and others. Toward reform of program evaluation: aims, methods and institutional arrangements. San Francisco: Jossey-Bass, 1980.

Cronbach, L.J. Designing evaluations of educational and social programs. San Francisco: Jossey-Bass, 1982.

De Bruyne, P., Herman, J., De Schoutheete, M. Dynamique de la recherche en sciences sociales. Paris: Presses universitaires de France, 1974.

Guba, E. Criteria for assessing the trustworthiness of naturalistic inquiries. Educational communication and technology, vol. 29, N° 2, Summer 1981, p. 75-91.

Nadeau, M.-A. L'évaluation des programmes d'études: théorie et pratique. Québec: Les Presses de l'Université Laval, 1981.

Scriven, M. The methodology of evaluation. In: R.E. Stake and others (Eds.). Perspectives on curriculum evaluation. AERA Monograph Series on Curriculum Evaluation, N° 1, Chicago: Rand McNally, 1967, p. 39-83.

Wilensky, H. Organizational intelligence: knowledge and policy in government and industry. New York: Basic Books, 1967.

Wolins, L. Research mistakes in the social and behavioral sciences. Ames, Iowa: The Iowa State University Press, 1982.

CONCEPTION FONCTIONNELLE

Une réponse fonctionnelle est possible, au problème du choix de la méthode.
Si l'on analyse bien la demande, on doit pouvoir savoir si le problème est :

- 1) de prouver une conclusion (de façon généralisable à un ensemble de situations).
- 2) de comprendre en profondeur une situation particulière (avec tous ses niveaux de signification).
- 3) de négocier une représentation commune à tous les groupes concernés (par un effort de communication systématique).

L'analyse de la demande doit aussi permettre de dire si l'évaluation se rapporte au passé, au présent ou au futur.

TYPE D'ETUDE ENVISAGE

FINALITE		PROUVER (Contrôle)	NEGOCIER (Cont. + Eval.)	COMPRENDRE (Evaluation)	OBJET D'ETUDE
FONCTION REGULATIVE	(Passé) Evaluation BILAN	Expérience de laboratoire	Pondération d'aspects objectifs et subjectifs	Jugement historique "situé"	Produit final
	(Présent) Evaluation FORMATIVE	Contrôle de conformité au déroulement prévu	Observation du processus et de ses déterminants extérieurs	Surveillance des changements d'opinion (par enquêtes)	Résultat intermé- diaire Processus
	(Futur) Evaluation PREDICTIVE	Pronostic à partir d'un modèle de simulation	Etude prospective	Consultation des intéressés sur leurs intentions	Causes Raisons Effets attendus

MOYENS UTILISES

EPISTEMOLOGIE	Evaluation externe	Evaluation négociée	Evaluation interne	Information à recueillir	Choix d'instruments
Sciences Naturelles				Faits Mesures	Indices statist. Tests objectifs Grilles d'obs.
Sciences Compréhensives				- Représen- tations sociales - Stratégies	Observations Questionnaires Interviews
Sciences Humaines				Perceptions Significa- tions Opinions	Interviews Modes d'expression divers