

L'EXPÉRIMENTATION PÉDAGOGIQUE A L'ÉCHELLE INTERNATIONALE

G. DE LANDSHEERE

en collaboration avec T.N. POSTLETHWAITE (1)

A cause de la complexité vertigineusement accrue de notre civilisation, les décisions de gestions, au niveau des entreprises comme au niveau des nations, ne se prennent plus sans études techniques préalables. Un mouvement similaire se dessine dans le domaine de l'éducation où le politique comme l'administrateur demandent de plus en plus souvent aux chercheurs des données objectives sur lesquelles ils puissent fonder leurs décisions. Elles concernent presque toujours des problèmes de très grandes dimensions.

Les renseignements qu'ils reçoivent en réponse ne sont pas toujours ni très utiles, ni très sûrs. Tantôt, ils proviennent d'une recherche à grande échelle, mais unique et cette unicité en rend les résultats contestables, tantôt ils sont communiqués sous forme de comptes rendus de plusieurs petites études aboutissant souvent à des résultats contradictoires. Les responsables ne sont donc pas ainsi très avancés et doivent bien se résoudre à prendre des décisions, parfois capitales, en se fiant à leur bon sens et à leur flair ou, plus prosaïquement, à leur réalisme politique.

Remarquons d'ailleurs tout de suite que, même si les résultats de la recherche étaient probants, le chercheur ne pourrait pas se substituer au politique à qui il appartient d'évaluer les implications sociales, économiques et culturelles des réformes introduites. En réalité, chercheurs en éducation et responsables politiques devraient être engagés en un dialogue permanent, et le bénéfice serait loin d'être unilatéral : le chercheur peut aider le politique à mieux poser les problèmes et à réunir des informations objectives, et le politique peut

(1) L'essentiel des considérations sur les surveys internationaux trouvent leur source dans une étude plus vaste due à T.N. POSTLETHWAITE : *International Surveys in Education*.

G. DE LANDSHEERE assume seul la responsabilité des propos concernant l'A.I.-P.E.L.F.

aider le chercheur à interpréter en termes plus sociaux les données qu'il a rassemblées.

Pourquoi l'expérimentation pédagogique internationale prend-elle une importance particulière dans ce contexte ?

L'expérience la plus courante comporte un groupe expérimental et un groupe de contrôle qui sert à évaluer les conséquences de la manipulation des variables. Mais, en éducation, tous les problèmes ne se prêtent pas à cette démarche (Carroll, 1963; Kish, 1965).

Par exemple, dans une expérience qui a pour objectif de comparer le rendement des classes homogènes et des classes hétérogènes, il importe que les professeurs enseignant dans un des deux systèmes croient à son efficacité. Comment remplir à coup sûr pareille condition ?

En fait, il est souvent difficile, sinon impossible, d'éliminer d'importantes variables perturbatrices de l'expérience. Ce qui explique pourquoi les résultats expérimentaux sont si souvent contradictoires.

Par ailleurs, il est malaisé de travailler sur des échantillons sûrement représentatifs des populations concernées par les décisions politiques, par exemple tous les enfants entrant dans l'enseignement secondaire.

Enfin, le *cadre naturel* dans lequel presque toutes les expériences pédagogiques devraient se placer, par exemple de véritables classes homogénéisées ou non homogénéisées, n'est pas souvent présent.

A ces égards, grande est la supériorité du *survey*, étude extensive dans un ensemble complexe considéré à un moment donné. Le *survey* porte sur « des échantillons aléatoires représentatifs, permettant de tirer des conclusions relatives à des populations bien définies et d'effectuer des mesures dans le cadre naturel des populations réelles » (Kish). Ici, les changements de situation ou de traitement se produisent spontanément, mais peuvent être considérés comme des manipulations expérimentales.

Les pays anglo-saxons surtout ont produit ces derniers temps des *surveys* nationaux de grande importance : Flanagan (1964), Coleman (1966), *Early Leaving* (1954), Crowther (1959), Newson (1964), Plowden (1967), Douglas (1964, 1968), etc.

Mais ces *surveys* nationaux connaissent eux aussi d'importantes limitations. Même dans un pays aussi décentralisé que les États-Unis où les écoles varient tant selon les régions, l'âge de scolarisation et la structure du système scolaire sont pratiquement partout les mêmes. Que dire alors de systèmes hautement centralisés comme en Belgique ou en France : à bien des égards, la variabilité est presque nulle. La recherche internationale permet de surmonter ces obstacles.

A condition de rassembler un nombre suffisant de pays, on peut disposer d'une palette comportant presque toutes les différences

possibles dans les systèmes scolaires, les programmes, les méthodes d'enseignement, la formation des maîtres, le statut socio-économique ou socio-culturel des familles, etc. Et, dans ces contextes naturels, les effets des différences sont très sensibles.

C'est pourquoi des *surveys* internationaux comme ceux de l'I.E.A. (Association Internationale pour l'Évaluation du Rendement Scolaire) revêtent une importance particulière (Postlethwaite et De Landsheere, 1969) :

1° Soit pour étudier la relation entre certaines variations dans l'action éducative et les apprentissages.

2° Soit pour contrôler la généralité ou l'universalité d'une relation découverte dans un pays.

L'I.E.A. a organisé, à l'heure actuelle, la plus grande série de *surveys* internationaux en éducation. Elle constitue donc un exemple privilégié.

Problèmes liés aux recherches pédagogiques internationales.

L'I.E.A. a d'abord essayé de décrire, d'une manière aussi détaillée que possible, les facteurs et les syndromes de facteurs liés au rendement scolaire. Ensuite, elle a entrepris de mesurer l'incidence différente de ces facteurs selon les pays et les contextes socio-culturels.

Les efforts ont d'abord été concentrés sur le domaine cognitif, parce que l'expérience acquise en matière de testing à grande échelle y était plus grande que dans le domaine affectif (qui n'a pourtant pas été tout à fait négligé). On a essayé de distinguer les facteurs qui interviennent de façon spécifique sur le rendement dans une branche (laboratoire de langues, exercices pratiques en sciences) de ceux qui influencent le rendement en général.

Semblable recherche est totalement étrangère à une sorte de compétition internationale. Car des comparaisons sommaires, qui ne tiendraient pas compte de programmes, du taux de rétention scolaire, etc. induiraient gravement en erreur. Il est, par exemple, peu sensé de comparer, à la fin de l'enseignement secondaire, le rendement moyen d'un pays où 10 % seulement du groupe d'âge se trouvent encore à l'école, après une sélection sévère, au rendement d'un autre pays qui retient 70 % du même groupe d'âge. Certains ajustements permettent cependant d'utiles parallèles.

En fait, dans un *survey* international, on procède en deux temps. On examine d'abord, à l'intérieur de différents pays, les relations qui peuvent exister entre divers aspects pédagogiques et les résultats obtenus par des élèves considérés isolément ou en groupes (par exemple, une école entière). Ensuite, on examine si des résultats semblables

correspondent à des pratiques semblables dans différents pays, ce qui équivaut à répéter une observation autant de fois qu'il y a de pays participants.

Les matières.

Jusqu'à présent, le choix des matières n'a guère soulevé de difficultés. La recherche sur le rendement en mathématiques dans douze pays, à 11, 14 et en dernière année du secondaire a servi de pierre de touche. Les sciences, la lecture, l'analyse textuelle, l'anglais et le français, enseignés comme langues étrangères ont suivi. On va sans doute s'attaquer maintenant à des recherches plus cliniques : éducation pré-scolaire, comportement des maîtres, processus de la lecture, stratégies de l'apprentissage.

Les variables.

Établir un cadre conceptuel soulève de plus grandes difficultés. En effet, pour définir les facteurs et les groupes de facteurs liés au rendement, dans différents systèmes scolaires, il faut tenir compte non seulement des facteurs pédagogiques, mais aussi familiaux, sociaux, économiques, etc. On se trouve ainsi devant des milliers d'influences éventuelles.

Pour opérer un choix, on procède en trois temps :

1^o On tire d'abord les enseignements de l'expérience acquise, aux niveaux national et international, par les chercheurs réunis. Une première série d'hypothèses est ainsi formulée.

2^o Des commissions nationales et internationales *ad hoc* sont constituées. Toute hypothèse doit être accompagnée de suggestions concernant les instruments permettant de l'éprouver.

3^o Enfin, des experts dans les différentes sciences sociales sont invités à signaler les facteurs qui, à leur avis, influencent le rendement scolaire.

Ayant procédé de cette façon, l'I.E.A. a actuellement retenu environ 250 variables pour les analyses des élèves, étudiés dans leur école, et 800 variables pour les analyses inter-écoles. Il est difficile d'aller au-delà dans le temps permis par les budgets de recherche et dans les conditions actuelles d'analyse statistique.

L'échantillonnage.

La définition des échantillons de populations à examiner nécessite aussi des études rigoureuses et de longues discussions. Le mérite exceptionnel de l'échantillon aléatoire est que l'erreur standard de

l'échantillon, considéré en tout ou en partie, peut être déterminée à partir des éléments intrinsèques à cet échantillon. Force est de constater que beaucoup de chercheurs ne dominent pas encore les techniques d'échantillonnage stratifié aléatoire; il faut donc les aider.

En gros, chaque population à examiner peut être divisée en une population échantillonnée et une population exclue (par exemple, l'enseignement spécial), à condition que l'absence de cette dernière n'entraîne pas une modification significative des résultats.

Dans la plupart des cas, on échantillonne en deux temps, d'abord par écoles ou par classes entières, puis par élève dans les écoles choisies.

Les statistiques nationales, y compris celles de la Belgique, réservent bien des difficultés aux chercheurs. Il n'est pas rare qu'il faille compenser les carences par mille démarches dans les ministères et ailleurs, et par des supputations inévitablement hasardeuses. Pour contrôler la représentativité des échantillons proposés par les différents pays, on inclut dans les données à réunir quelques variables-tests, pour lesquelles on dispose à l'avance de statistiques aussi sûres que possible...

La construction des instruments.

En principe, la construction d'instruments internationaux, — tests, échelles, questionnaires —, ne diffère pas méthodologiquement de la construction d'instruments nationaux. Mais l'ampleur des problèmes est tout autre.

Le canevas des tests de connaissances résulte en bonne partie d'une analyse du contenu des programmes scolaires. Or, plus on s'élève dans le système scolaire, moins les programmes des différents pays ont des points communs.

On adopte la démarche suivante. On prend d'abord en considération tous les domaines communs, puis d'autres que les experts estiment devoir retenir.

Pour permettre néanmoins une comparaison équitable des résultats obtenus, on pondère les résultats partiels sur la base d'une analyse des composants principaux.

Il est inutile d'insister ici sur l'importance capitale du prétest dans une entreprise internationale que la traduction en plusieurs langues d'un même ensemble d'instruments complique comme à plaisir.

Dans la recherche sur les mathématiques, on demandait, par exemple, aux chefs d'établissements d'indiquer si leur école était du type « comprehensive » ou non. Les analyses ont montré que ce mot a été compris très différemment selon les pays.

Traitement des données.

Le plan général du traitement des données doit évidemment être dressé très tôt. Toutefois, les techniques liées à l'ordinateur évoluent si vite qu'il est sage de conserver autant de souplesse que possible. On arrive rapidement à des volumes d'information considérables. Testant quelque 350 000 étudiants dans une vingtaine de pays, l'I.E.A. se trouve actuellement devant environ 300 millions d'informations à traiter...

Pour le traitement et les opérations d'administration, le *dry-run* ou essai général à petite échelle est tout aussi important que le pré-testing dans la construction des instruments.

Aux différents stades du traitement des données, il importe de procéder à des contrôles rigoureux. Aussi, dès que les statistiques univariées sont obtenues pour une population d'un pays, on les lui envoie pour vérification. Par la même occasion et à titre de bonne relation publique, on envoie à chaque école participante les scores obtenus par chaque étudiant, les scores totaux ou partiels, les moyennes et les écarts types de l'école, ainsi que les données d'ensemble pour chaque type d'écoles du pays.

Dans les analyses statistiques, les scores obtenus aux tests et aux échelles d'attitudes sont, en général, considérés comme les résultats de l'éducation (variables dépendantes), les données apportées par les questionnaires constituent les variables d'entrée et de traitement (variables indépendantes). Il n'est pas toujours facile de décider si l'on se trouve devant des variables « d'entrée », « de traitement » ou « de sortie ». Prenons l'exemple d'un élève examiné à l'âge de 14 ans. A cet âge-là, il est acquis, qu'il est né, a été élevé par ses parents dans un certain entourage et a déjà fréquenté l'école longtemps! Le *survey* va simultanément porter sur certains aspects de ce passé et sur les études présentes, ainsi que sur les attitudes et le rendement actuels. Un des buts est de découvrir dans quelle proportion les facteurs présents et passés expliquent les différences de rendement et d'attitudes. En d'autres termes, si le rendement et les attitudes sont considérés comme réponses, et les autres facteurs comme conditions, on essaie de déterminer dans quelle proportion les conditions expliquent les réponses. Deux complications surgissent. En premier lieu, une réponse peut dépendre, non pas simplement d'un groupe de conditions, mais de toutes les conditions à la fois. En second lieu, il n'est pas facile de savoir si les ensembles de conditions agissent simultanément ou l'un après l'autre. Enfin, il est utile de distinguer, aussi clairement que possible, les conditions dépendant du système éducatif, de celles qui lui sont extérieures.

Une façon systématique de procéder serait de commencer par

évaluer l'effet d'une seule condition, sur une réponse, puis de voir comment cette estimation se modifie si l'on tient compte des effets des autres conditions appartenant au même groupe et, finalement, de voir quelles autres modifications interviennent encore si l'on tient aussi compte des effets des autres groupes de conditions. Ceci signifie techniquement qu'il faut obtenir une série d'équations de régression, d'abord pour la réponse à une condition, puis à un ensemble de conditions, et finalement à tous les ensembles de conditions. Comme les conditions sont nombreuses, des équations plus nombreuses encore sont nécessaires. On peut cependant diminuer ce nombre en considérant que certaines conditions agissent avant d'autres, dans le temps.

On vient de faire plusieurs fois allusion à des ensembles de conditions. Il n'est pas toujours aisé de constituer ces groupes, surtout lorsque les degrés de liberté sont peu nombreux, comme c'est par exemple le cas pour les analyses inter-écoles. Mais ce problème, quoique fascinant, n'est pas particulier à la recherche pédagogique internationale.

Problèmes d'organisation et d'administration.

On ne peut sous-estimer la complexité des problèmes d'organisation posés par la recherche internationale en éducation.

D'abord parce que le niveau de qualification des chercheurs varie considérablement. J'ajoute immédiatement que le perfectionnement résultant de la participation à une recherche internationale de haut niveau constitue, à mes yeux, un des principaux bénéfices de cette participation.

Ensuite parce que les communications sont lentes et compliquées. Non seulement à cause de la lourdeur du travail, mais aussi parce que le peu de chercheurs de haute qualification existant actuellement dans nos pays sont souvent débordés et manquent d'aide. La langue de communication n'est d'ailleurs pas si facile à choisir, sauf dans une association comme le nôtre.

Une coordination internationale est nécessaire et elle coûte cher. L'I.E.A. travaille sur trois continents et a donc des frais plus grands que ne le seraient ceux pour les pays de langue française. Il n'est cependant pas inutile de retenir que, de 1969 à 1972, les frais internationaux de l'I.E.A. s'élèvent à environ 5,5 millions de nouveaux francs français...

Du survey à l'expérimentation.

Au fur et à mesure que les analyses des données recueillies par *surveys* s'affineront, on passera des études corrélationnelles aux expli-

cations causales, avec une certitude croissante. Des expériences proprement dites devront alors remettre les conclusions à l'épreuve, car en fin de compte seule l'expérimentation peut établir rigoureusement un déterminisme.

A mi-chemin entre le *survey* classique et l'expérience sur le terrain, se trouve le testage de modèle théorique par *surveys*.

Par exemple, Beeby (1966) a proposé une théorie des stades de développement de l'école primaire qui permet notamment d'établir une relation entre le style d'enseignement et le niveau de développement économique d'un pays. Pareil modèle peut très probablement être vérifié en étudiant un éventail de pays suffisamment contrastés au point de vue de la variable indépendante.

D'autres modèles attendent d'être testés.

Noah et Eckstein (1969) ont articulé un ensemble d'hypothèses sur les relations entre certaines dimensions de l'éducation et certaines caractéristiques des structures sociales, économiques, religieuses ou politiques ou, encore, les idéologies sociales existant dans une communauté.

Noonen (1970) a, de son côté, proposé un modèle pour étudier le degré d'égalité des chances en éducation.

Mais les *surveys* internationaux peuvent plus encore qu'éprouver des modèles, ils peuvent en faire naître. Grâce, en particulier, aux nouvelles techniques d'analyse de pistes causales, des modèles de cause à effet peuvent, par exemple, être construits pour étudier l'influence des origines sociales, des maîtres, des variables scolaires, sur le rendement scolaire.

L'action de l'A.I.P.E.L.F.

Les chercheurs de langue française doivent-ils rester indifférents à ce nouveau type de recherche que l'avènement de l'informatique ne rend d'ailleurs possible que depuis moins de dix ans? Évidemment non et plusieurs d'entre nous ont déjà joué un rôle important dans l'I.E.A.

Personne, je crois, ne songerait à contester les mérites exceptionnels de l'I.E.A. qui a indiscutablement ajouté une dimension nouvelle à la recherche en éducation et est aujourd'hui le plus grand organisme mondial de recherche pédagogique.

Mais cette dimension mondiale et l'hégémonie bien explicable des chercheurs anglo-saxons ne permettent peut-être pas de cerner nos problèmes aussi finement que nous le souhaiterions et ne tiennent pas assez compte de nos caractéristiques culturelles.

Occasionnellement déjà, des collaborations entre S. ROLLER, F. HOTYAT, G. MIALARET, moi-même et d'autres membres de notre

association sont nées. Le moment n'est-il pas venu de lancer aussi certains projets propres aux pays de langue française? Ils devaient d'abord être très limités, car même là où l'expérience de la conduite des *surveys* internationaux est déjà acquise, les disponibilités humaines ne sont pas grandes. Je crois néanmoins que nous avons tout ce qu'il faut pour commencer, sauf l'argent nécessaire au fonctionnement du centre de coordination. On doit pouvoir le trouver.

Pourquoi les pays de langue française seraient-ils les parents pauvres de la recherche en éducation?