

« Le double aveugle contre Pygmalion »  
Éléments de psychosociologie de la recherche en éducation et  
méthodologie des plans

Jean-Marie Van der Maren

Volume 3, Number 3, Fall 1977

URI: <https://id.erudit.org/iderudit/900057ar>

DOI: <https://doi.org/10.7202/900057ar>

[See table of contents](#)

Publisher(s)

Revue des sciences de l'éducation

ISSN

0318-479X (print)

1705-0065 (digital)

[Explore this journal](#)

Cite this article

Van der Maren, J.-M. (1977). « Le double aveugle contre Pygmalion » : éléments de psychosociologie de la recherche en éducation et méthodologie des plans. *Revue des sciences de l'éducation*, 3(3), 365–380. <https://doi.org/10.7202/900057ar>

Article abstract

Lorsqu'un chercheur en éducation entreprend une recherche qui implique des sujets (élèves, administrateurs, professeurs, expérimentateurs,...), il instaure un réseau de relations qui n'est pas sans influencer les résultats de ses travaux. À partir de certaines situations de recherche, cet article envisage les différents types de contaminations qui originent de ces relations : que ce soit dans le recueil des données (stéréotypie, halo, etc...), dans la production des performances (exaltation, consignes implicites, défense, effet Rosenthal, etc...), ou dans leur combinaison (effet Pygmalion). Ensuite, à l'occasion de l'analyse d'une recherche psycho-pharmacologique réalisée en milieu éducatif, se dégage un plan d'expérience (à 6 groupes) qui, combiné à la stratégie du double aveugle, a pour fin de contrôler les contaminations d'abord envisagées.

# “Le double aveugle contre Pygmalion”

Jean-Marie Van der Maren \*

## ÉLÉMENTS DE PSYCHOSOCIOLOGIE DE LA RECHERCHE EN ÉDUCATION ET MÉTHODOLOGIE DES PLANS

### RÉSUMÉ

Lorsqu'un chercheur en éducation entreprend une recherche qui implique des sujets (élèves, administrateurs, professeurs, expérimentateurs,...), il instaure un réseau de relations qui n'est pas sans influencer les résultats de ses travaux. À partir de certaines situations de recherche, cet article envisage les différents types de contaminations qui originent de ces relations : que ce soit dans le recueil des données (stéréotypie, halo, etc...), dans la production des performances (exaltation, consignes implicites, défense, effet Rosenthal, etc...), ou dans leur combinaison (effet Pygmalion). Ensuite, à l'occasion de l'analyse d'une recherche psycho-pharmacologique réalisée en milieu éducatif, se dégage un plan d'expérience (à 6 groupes) qui, combiné à la stratégie du double aveugle, a pour fin de contrôler les contaminations d'abord envisagées.

Bien des efforts ont été déployés ces dernières années afin de donner un statut scientifique autonome aux recherches entreprises dans les sciences de l'éducation. On y a vu l'appropriation d'un certain nombre de techniques statistiques, de schémas expérimentaux, puis leur critique et leur adaptation aux conditions particulières du champ de l'éducation. Mais tout en découvrant ce qui pouvait peut-être faire l'originalité de certains problèmes de recherche en éducation, il semble que plus d'un chercheur en ce

---

\* Van der Maren, Jean-Marie : professeur, Université de Montréal.

domaine soit encore sourd à certains travaux méthodologiques qui, par ailleurs, ébranlaient la conscience des praticiens.

En effet, si dans certains cercles académiques les travaux de Rosenthal, plus particulièrement le livre écrit en collaboration avec Jacobson (1968), ont suscité maintes discussions, ils semblent n'avoir eu que peu d'effet sur l'élaboration des schémas de recherche. Ils ont, apparemment, été pris au pied de la lettre et embarqués au creux du débat anti-école avec le flux des analyses sociologico-gauchistes à propos de la fonction et du rôle de l'institution scolaire dans nos sociétés. Pourtant, les publications de Rosenthal avaient une autre visée : dévoiler les biais méthodologiques de la situation expérimentale dus aux attentes, aux anticipations des expérimentateurs. Et si le psychologue social peut être suspecté d'influencer les performances de ses sujets par le fait même qu'il a des hypothèses, n'est-on pas en droit d'être encore plus suspicieux à l'égard du pédagogue ? N'est-il pas souvent plus engagé que le psychologue dans la défense de ses hypothèses ? Lorsqu'il s'agit d'évaluation de programmes, de méthodes ou d'outils pédagogiques, ne devrait-on pas manifester les mêmes exigences que lorsqu'il est question d'évaluer les effets d'une médication nouvelle ?

Certains projets de recherche auxquels nous avons été affrontés ont attiré notre attention à la fois sur la bonne foi évidente de pédagogues-chercheurs et en même temps sur leur négligence de l'influence possible, et quasi certaine, de l'expérimentateur sur les résultats de ses travaux. D'ailleurs, ils sont bien souvent si sûrs de la qualité de leur démarche pédagogique qu'ils sont étonnés qu'on leur dise que ce ne sont là que des prophéties qui s'exaucent d'elles-mêmes, et que leurs efforts « d'expérimentation » ressemblent étrangement à ceux du pompier occupé d'enfoncer une porte ouverte (encore ce dernier a-t-il l'excuse de n'y rien voir dans la fumée). Cependant, les avertissements de Rosenthal étaient clairs, étayés, diffusés dès 1964, et présentés en français par Lemaine en 1969.

L'examen d'un projet de recherche en éducation illustrera cette situation. Il s'agissait d'un travail, présenté par deux étudiants de 2<sup>e</sup> cycle, qui vise à prouver l'effet de l'entraînement au niveau d'un présumé « pré-requis » sur le développement qualitatif d'un acquis académique classique. Ils définissent avec précision les paramètres de l'acquis académique à mesurer, prévoient un pré- et un post-test de cet acquis encadrant l'entraînement au pré-requis. De plus, ils décident de procéder avec groupe expérimental et groupe contrôle. Mais, première remarque, ils ne prévoient pas de mesure pour ce pré-requis ! Ainsi ils tentent de prouver l'existence d'un lien entre une variable dépendante (acquis) et une variable indépendante (pré-requis) sans évaluer les variations de cette dernière. D'une part, ce n'est pas parce que les chercheurs sont eux-mêmes responsables de l'entraînement au niveau du pré-requis qu'ils peuvent être sûrs de son effet ; de plus, sans mesure, ils ne peuvent en connaître l'ampleur. Mais d'autre part, ils ne peuvent non plus avancer quoi que ce soit quant à la maîtrise de ce pré-requis dans le groupe contrôle (qui ne reçoit pas l'entraînement) tant au moment du pré-test que du

post-test. Dès lors, la vraie variable indépendante n'est plus le pré-requis, mais seulement leur présence ou leur absence en tant qu'entraîneur.

Ce premier défaut méthodologique, assez facile à corriger quand on s'en rend compte à temps, indique assez quelle peut être l'implication du pédagogue-chercheur, et attire d'autant plus l'attention sur le risque d'une influence au niveau des résultats.

Si l'on poursuit l'examen du projet, on constate que ce risque est bien grand. La recherche s'effectue dans deux milieux scolaires simultanément (aspect socio-économique). Dans chacune des écoles il y aura donc groupe expérimental et groupe témoin, pré-test, entraînement et post-test. Pré-test et post-test seront appliqués par le chercheur qui, dans l'école envisagée, n'est pas le responsable de l'entraînement au niveau du pré-requis. De plus, les résultats au pré-test ne sont pas communiqués à celui qui entraîne. Mais les groupes sont constitués a priori par les écoles (sur quelle base ?) et les épreuves seront appliquées en alternance, enfant témoin, enfant expérimental. En clair, cela signifie que le testeur, quel qu'il soit, saura, aux moments des pré- et post-tests, quel enfant recevra et aura reçu l'entraînement, ainsi que ceux qui ne le recevront pas. Il est clair dès lors qu'aucune garantie n'est présente contre un effet Pygmalion. Évidemment, le chercheur responsable de l'entraînement ignore les résultats au pré-test, et ne peut donc modifier ses attitudes pédagogiques en fonction de ces résultats. Mais ici l'importance d'une telle modification est minimisée puisqu'il n'a aucun contact avec les enfants du groupe témoin. Par contre, le testeur risque d'être plus qu'influencé : c'est lui qui aurait dû être dans l'ignorance de qui est membre du groupe expérimental ou du groupe témoin. En conséquence dans un tel plan en se partageant les rôles et les ignorances, les chercheurs jouent mal à l'aveugle : ils ne sont que borgnes et verront quand même leurs résultats influencés par leurs hypothèses. Car même avec des épreuves ayant toutes les qualités psychométriques voulues, c'est dans une relation, expérimentateur-sujet, que s'effectuent les pré- et post-tests. Dès lors le testeur, connaissant l'appartenance des sujets à l'un ou l'autre des groupes, aura bien des difficultés à ne pas influencer les résultats de ceux-ci dans le sens de ses attentes. Les travaux de Rosenthal sont assez probants à ce sujet : même si des erreurs d'observation pourront, peut-être, être évitées, les réponses des sujets, de par les modes d'interaction qui se déploieront, seront plus que probablement influencées dans le sens des hypothèses mises à l'épreuve.

Encore que l'honnêteté foncière des expérimentateurs ne les mette pas à l'abri d'erreurs d'observation. Une telle mésaventure a été récemment vécue par deux collègues alors que l'objet de leurs observations n'avait guère de lien avec leurs hypothèses. Voulant donner une base plus quantitative aux classements effectués par des enseignants à propos d'un apprentissage scolaire classique, ils avaient construit une épreuve telle que la somme des réussites et des erreurs devait toujours aboutir au même nombre. Quel ne fut pas leur étonnement de découvrir, lors d'un contrôle qu'ils pensaient être routinier (l'établissement de la somme des réussites et des erreurs), que systématiquement ces sommes n'étaient pas exactes pour un certain nombre de « bons » élèves : ou ils n'avaient pas observé, ou ils n'avaient pas compté les erreurs dans les protocoles des

bons élèves ! Il leur fallut recommencer toutes les corrections, heureux cependant de disposer d'une épreuve où tel biais pouvait être dépisté.

Ces deux cas en illustrent bien d'autres et soulignent la nécessité pour les chercheurs en éducation d'être attentifs à une série de contaminations possibles de leurs résultats, et de mettre en place des procédures du type « double aveugle » (ou double insu) ou des plans avec « groupe témoin à anticipation » (avec pseudo-traitement) ou encore une combinaison des deux.

Quelles sont les contaminations possibles des observations ou des résultats dus aux attentes de l'expérimentateur ou même à la seule situation expérimentale ?

Sans être exhaustif, soulignons quatre types de contaminations \* *dans le recueil des données* qui peuvent être assez fréquents en éducation.

Il y a d'abord l'effet de stéréotypie : L'observateur ou l'examineur classe, plus ou moins consciemment, le sujet dans une catégorie à laquelle sont habituellement reliés certains traits. Ce premier classement peut s'être fait soit à partir d'une information préalable, soit à partir de l'agglomération des premières perceptions. À la suite de quoi (et c'est là que se manifeste l'effet), l'examineur peut avoir tendance à ne plus percevoir chez ce sujet que les traits reliés à sa catégorie et à scotomiser les traits qui ne correspondraient pas au classement. Un exemple en a déjà été rapporté : les examinateurs qui avaient omis de compter les erreurs dans les protocoles des « bons élèves ».

Le classique effet de halo tend à s'en rapprocher : c'est la tendance à relier certains faits d'observation subséquents à une première série d'indices observés sans qu'un lien objectif ne relie les deux séries d'éléments observés. Mais le lien subjectivement établi entre les deux séries n'a d'autre fonction que de renforcer la première impression ou le premier jugement de l'observateur. Ainsi, un superviseur visitant un stagiaire qu'il estime empathique au bout de trois à cinq minutes risque d'avoir tendance à compter comme empathiques des comportements qui n'ont rien à voir avec ce concept rogérien : toutes les « bonnes » attitudes seront mises au compte de l'empathie, et le stagiaire sera étiqueté comme vraiment empathique.

Autre contamination de l'observation : la propension que peut avoir l'observateur à noter plus ou moins d'événements en fonction de ses hypothèses. Il y a là hyper ou hypo-perception. Il peut en aller de même lorsqu'il s'agit de dépouiller de façon interprétative des protocoles (par exemple, dans une analyse de contenu). Ces erreurs de perception, d'évaluation ou d'interprétation peuvent par ailleurs aller dans les deux sens : certains observateurs favorisant leurs hypothèses, d'autres au contraire les défavorisant selon une tendance à se montrer « plus catholique que le pape ».

Enfin, certains chercheurs en éducation pourraient avoir tendance à protéger leurs hypothèses en ne recueillant pas certaines catégories d'éléments ou de faits

---

\* Bien que le terme « biais » soit classiquement utilisé pour recouvrir l'ensemble de ces problèmes, nous préférons « contamination » qui n'a pas la connotation péjorative de « biais ».

pertinents qui risqueraient de « brouiller les cartes », de rendre les résultats moins évidents. C'est un phénomène de perception sélective. Ainsi, certains peuvent ne prendre en considération que certaines dimensions du phénomène étudié et simplifier en conséquence le champ théorique où s'insèrent les hypothèses. Un exemple courant en éducation est le mépris d'événements relevant d'effets appelés « secondaires ». Peut-être secondaire pour le chercheur, mais non pas pour la réalité scolaire. Dans certains cas l'élimination de dimensions à observer est une nécessité méthodologique (lourdeur de l'instrumentation), mais dans d'autres cas cela ressemble fort à la technique du vendeur qui pour mieux vendre son produit préfère en ignorer les défauts.

Soulignons encore une fois que ces contaminations de l'observation, tout comme les contaminations des résultats dont on parlera plus loin, peuvent être et sont même souvent inconscientes et manifestées par des chercheurs dont on ne peut mettre en doute l'honnêteté. Il s'agirait seulement, lorsqu'on connaît ces risques bien involontaires, de prévoir des méthodes permettant sinon de les éliminer, au moins de les contrôler, de les mesurer.

Les *contaminations des résultats* sont bien connues en psychologie sociale et peuvent tout autant se manifester en éducation.

Envisageons d'abord les *contaminations des résultats dues aux sujets*. Le simple fait qu'ils se rendent compte de leur participation à une expérience va probablement modifier leurs comportements. Bien que n'étant ni des volontaires ni des étudiants en psychologie qui tentent de se comporter en « bons sujets » (c.à-d. de manière à ce que les hypothèses de départ se trouvent confirmées), les élèves sont sensibles à toute expérimentation les concernant. Cela d'autant plus qu'il est bien rare qu'on puisse mener une quelconque recherche sur le terrain en éducation sans modifier l'environnement (classe, personnes, observateurs,...) ou la structure des conduites scolaires (horaires, manières de procéder, etc...). Dès lors, après un premier temps d'observation pendant lequel les élèves vont tenter de découvrir le sens de ce qui se passe autour d'eux, cette découverte va les amener à réagir de quatre manières différentes, ces réactions modifiant les résultats.

Un premier type de réaction est une sorte d'exaltation à la suite de la découverte de leur importance, puisqu'on expérimente chez eux. Cette exaltation peut avoir des effets allant aussi bien dans le sens des hypothèses que contre lui, ou encore n'ayant rien à voir avec celles-ci mais brouillant toute observation les concernant.

Au contraire, une autre réaction typique de certaines écoles proches des centres universitaires, ou cibles privilégiées des chercheurs étant donné certaines de leurs caractéristiques, est une forme d'apathie. Ces élèves en ont « assez » de voir leur environnement et leurs modes de fonctionner sans cesse bousculés, et en arrivent à souhaiter ne plus être sujets d'expérience afin de bénéficier d'un enseignement normal (...mais il y a beaucoup de chance pour qu'on ne demande pas leur avis !). Comme

dans le cas de l'exaltation, les effets de cette attitude peuvent rendre les résultats hyper ou hypo-concordants avec les hypothèses, ou encore tout neutraliser.

Dans ces deux premiers types de réaction, se manifeste une surdité à certains éléments des consignes, car ce qui importe n'est pas le contenu de l'expérience mais le fait qu'on expérimente. Ainsi, par exemple, dans une école pour déficients mentaux, tant aux pré- et post-tests que pendant le traitement, les enfants avaient tendance à scotomiser l'aspect vitesse de façon à rester plus longtemps en présence des expérimentateurs qui venaient de l'université alors que cette tendance ne se manifesta pas lors d'une contre-expérience réalisée par l'intermédiaire du personnel habituel de l'établissement.

Les deux autres types de réactions sont liés à la perception par les élèves d'indices qui prennent pour eux valeur de consignes implicites soit quant aux hypothèses en jeu, soit quant à leur statut dans la recherche (par exemple, groupe expérimental vs. groupe témoin). En ce qui concerne la perception d'indices quant aux hypothèses, la réaction des élèves dépendra essentiellement de leur position affective à l'égard des expérimentateurs ; cela plus les élèves sont jeunes et plus ils sont socio-émotivement perturbés. Découvrant par les rumeurs qui circulent ou dans le déroulement de l'expérience des indices quant au sens de celle-ci, ils pourront être tentés soit de vouloir faire plaisir et dès lors de se comporter de manière à confirmer les hypothèses de départ (du moins ce qu'ils croient qu'elles sont), soit de faire « enrager » leur monde. Dans ce dernier sens, ils peuvent vouloir profiter de cette occasion pour régler le compte resté ouvert d'un conflit parfois ancien en se comportant comme de mauvais sujets, réagissant dans un sens opposé aux hypothèses perçues ou imaginées.

Enfin, la découverte d'indices quant à leur statut dans la recherche peut aussi amener les enfants à modifier leurs comportements et en conséquence à fausser les résultats. En effet, les enfants risquent de ne plus réagir de la même façon après la découverte de leur appartenance à un groupe témoin qui ne bénéficie pas du traitement offert au groupe expérimental, et inversement. Or, lorsque des modifications dans leur environnement ou dans les manières de les conduire interviennent, les enfants ont tendance à s'observer, à communiquer entre eux afin de savoir ce qui se passe ; ainsi ils risquent de découvrir que le traitement n'est pas identique. Leurs réactions peuvent aussi être modifiées s'ils en viennent à croire qu'ils ont été sélectionnés en fonction de certains traits de personnalité, de certaines de leurs caractéristiques académiques ou sociales. Lorsqu'enfin les élèves parviennent à combiner les deux séries d'indices (hypothèses et groupes), leurs comportements manifesteront une tendance à s'aligner sur ce qu'ils croient être attendu d'eux, ou au contraire à systématiquement s'en écarter. Quoique nous avons rencontré un cas où la modification du comportement, bien que liée à la perception de ces indices, n'était pas fonction de la relation expérimentateur-sujet. Il s'agissait d'un adolescent classé « caractériel », et plus particulièrement opposant, qui

profita d'une expérimentation pour changer son attitude, ce qui annula un pairage, et cela parce qu'il en avait « assez de faire partie de la bande des caves ».

Si les élèves interprètent les éléments de la situation expérimentale en fonction de leur rapport à l'expérimentateur ou au chercheur, un risque semblable est encore plus grand lorsque la recherche porte sur la pratique pédagogique des maîtres ou des administrateurs. Lorsque ceux-ci acceptent de faire partie d'une recherche, qu'en espèrent-ils en retour ? Quels éléments de leur relation au chercheur ont pu déclencher leur participation à l'expérience ? Et en conséquence, quelles attitudes vont-ils avoir à l'égard des consignes ?

Dans certains cas l'attitude est a priori positive et pourrait se rapprocher de l'exaltation que peuvent vivre les enfants. Participer à une telle expérience enfle les perceptions qu'ils se font d'eux-mêmes et les stimule parce que cela les rapproche de l'image de l'enseignant-chercheur, en « éducation permanente », soucieux du progrès pédagogique. Ils adhèrent à une telle image idéale soit depuis leur formation, soit depuis leur rencontre avec certains animateurs ou encore depuis leur participation à l'un ou l'autre « stage ». Dès lors participer à une recherche d'une part satisfait leur narcissisme en restreignant l'écart entre l'image de soi et l'image idéale de soi pour soi, et d'autre part les range dans une catégorie à part parmi leurs collègues : l'élite pédagogique des enseignants-chercheurs, des innovateurs, des progressistes !

D'autres, au contraire, risquent de réagir en fonction du dérangement que la recherche occasionne à leurs routines. Leur participation ne sera pas très active, minimale au niveau de leur implication personnelle, même s'ils sont capables de présenter de grandes démonstrations d'intérêt, jusqu'au moment où, un élément de la recherche les ayant dérangés, ils réagiront négativement, sabotant même le travail entrepris. Ainsi, lors d'une recherche projetée afin de vérifier certaines hypothèses quant aux phases de la négociation scolaire, l'administration d'une institution (avec internat) avait accepté l'ensemble du projet. Elle avait fourni du personnel (observateurs) pour deux mois, modifié les grilles horaires, consenti à \$2.250 en matériel d'enregistrement et d'observation et entrepris des travaux dans une classe afin d'y installer un faux mur avec miroir sans tain. Une dernière contrainte avait été acceptée : la population de la classe où l'expérience se déroulerait devait rester stable. Or, à mi-chemin de l'expérimentation (après 15 jours), cette population fut changée aux 4/10<sup>e</sup> parce que des lits étaient devenus inoccupés dans une section, et que pour les remplir il fallait faire tourner les élèves ! Malgré l'investissement, l'expérience bien commencée ne put jamais être complétée.

Dans d'autres cas, l'attitude est plus mitigée : on n'y croit pas trop, on ne s'emballe pas, mais on va se comporter au mieux afin d'avoir une bonne note de la part de la hiérarchie, afin de se ménager une « image » aux yeux des collègues comme des responsables de la recherche. Dès lors, si dans leur exaltation les premiers risquent de ne

pas comprendre ce qui est attendu d'eux en faisant de la recherche d'un autre leur propre recherche (par exemple, modification des consignes pour adapter la tâche à leurs propres objectifs avec leur groupe d'enfants), ceux-ci vont tout faire pour satisfaire les attentes du chercheur. Entre autres, ils vont essayer de deviner ce qu'ils pourraient faire, au-delà des consignes qu'on leur donne, pour présenter l'image d'un bon collaborateur, qui fait ce qu'on lui demande et qui peut même en faire plus s'il croit que cela peut faire plaisir (p.ex. donner un entraînement spécial aux élèves en dehors des séances d'observation afin d'améliorer les performances, etc...).

Par ailleurs, certains percevront les phases de la recherche comme autant de formes d'inquisition. Ils y participent, pour ménager une image ou un statut, mais ils maquillent la réalité, non pas pour satisfaire les chercheurs, mais pour protéger l'image qu'ils croient devoir présenter, soit d'eux-mêmes, soit de leurs élèves. Ainsi, lors d'une pré-enquête afin de mettre au point un système de mesure d'habiletés professionnelles dans une recherche visant la validation d'une batterie d'orientation, un groupe d'enseignants avait systématiquement remonté les notes attribuées aux élèves. Cela parce que, pour eux, un bon enseignant était quelqu'un chez qui tous les élèves réussissent. Aussi combien ne furent-ils pas dépités lorsque le chercheur confronta les notes attribuées sur les protocoles avec les notes effectivement inscrites dans les bulletins remis aux élèves.

Ces contaminations des résultats provoquées par les enseignants collaborant à une recherche indiquent la possibilité d'une semblable *contamination par le chercheur* lui-même, ce que Lemaine et Lemaine appellent l'effet Rosenthal, ce que d'autres ont appelé effet Pygmalion.

Après avoir noté dans de nombreuses anecdotes de lourdes présomptions quant à ce phénomène, Rosenthal mit en place une série de recherches tendant à démontrer son existence. Les résultats qu'il rapporte et ceux d'autres chercheurs ne peuvent plus laisser de doute : dans bien des cas les hypothèses que le chercheur tente de vérifier contaminent ses résultats. Ses attentes influencent la performance des sujets, qui dès lors comblent ses espoirs. On ne peut, à première vue, empêcher un chercheur de vouloir vérifier des hypothèses ; or, dès ce moment, il apparaît que ses attentes vont être communiquées — par des canaux variables et dont le jeu n'est guère précisé à ce jour — aux sujets qui réagiront à cette communication inconsciente et involontaire comme ils réagissent à l'ensemble des consignes et des éléments « significatifs » de la situation. En conséquence, un certain nombre de recherches risquent de n'être que des prophéties qui se réalisent d'elles-mêmes.

Ce phénomène est d'autant plus probable que, comme Rosenberg (in : Rosenthal et Rosnow) le signale, il semble exister une dynamique de la communication expérimentateur-sujet qui englobe plus que l'effet Rosenthal. Il y aurait une série d'interactions-ajustements entre : les attentes de l'expérimentateur, la recherche par les sujets de consignes implicites, et conséquemment l'élaboration d'hypothèses quant à la conduite à tenir, les réactions de l'expérimentateur aux conduites des sujets, ces réactions étant

perçues par les sujets comme des évaluations, nouveaux indices qui commandent un ajustement de leurs conduites. C'est par ailleurs une dynamique semblable qui règle normalement la communication entre maître et élèves dans les classes, chacun tentant d'ajuster sa conduite, son message à celui de l'autre. C'est là le véritable effet pygmalion tel que l'illustre la pièce de B. Shaw.

Afin de mieux examiner l'ampleur du problème méthodologique posé et les stratégies de contrôle possible, résumons les principaux types de biais ou d'artéfact jusqu'ici examinés. Nous les classons en trois catégories :

- A. Contaminations dans le recueil ou l'interprétation des données
  - A.1. stéréotypie
  - A.2. halo
  - A.3. Hyper- ou hypo-perception
  - A.4. perception sélective
  
- B. Contaminations dans la production des résultats
  - B.1. dues aux élèves
    - B.1.1. réactions au fait expérimental
      - B.1.1.a. exaltation
      - B.1.1.b. apathie
    - B.1.2. réactions aux données de la situation (consignes implicites)
      - B.1.2.a. par rapport aux hypothèses supposées
      - B.1.2.b. par rapport au statut du sujet (gr. exp. ou témoin)
  - B.2. dues aux maîtres et/ou aux administrateurs
    - B.2.1. réactions au fait expérimental
      - B.2.1.a. inflation de l'image et exaltation
      - B.2.1.b. protection de la routine et implication minimisée
    - B.2.2. réactions aux données de la situation
      - B.2.2.a. hyper-participation
      - B.2.2.b. défense contre l'évaluation
  - B.3. dues au chercheur : effet Rosenthal
  - B.4. dues à la conjonction de B.1., B.2. et B.3. : effet Pygmalion
  
- C. Contaminations et dans le recueil des données (A) et dans la production des résultats (B).

Sans faire un inventaire de l'ensemble des stratégies mises en avant ces vingt dernières années (voir par exemple Rosenthal et Rosnow), nous voulons insister sur la méthode dite du « double aveugle » ou encore du « double insu », bien que pour être correctement utilisée elle exige souvent plus de deux « aveugles ». Le principe de base en est simple : les sujets ne doivent pas savoir s'ils font partie du groupe expérimental ou du groupe témoin, s'ils reçoivent ou non le traitement ; l'observateur qui recueille les données tant au pré-test, au post-test que pendant le traitement ne doit pas savoir qui

parmi les sujets fait partie du groupe expérimental ou du groupe témoin. De plus, idéalement, l'expérimentateur qui applique le traitement ne doit pas savoir s'il applique le vrai ou le pseudo-traitement, ni quelles sont les hypothèses en question liées aux traitements. Aussi, comme nous allons l'illustrer à partir d'une étude pharmacologique réalisée en milieu orthopédagogique, cela aboutit souvent à faire beaucoup d'aveugles, chacun des acteurs engagés dans la recherche ignorant un petit quelque chose au sujet de l'ensemble de la recherche.

Afin de mieux saisir comment la méthode du double insu permet de neutraliser sinon d'éliminer certaines des contaminations envisagées, nous analyserons une recherche non publiée, n'ayant pas abouti aux résultats escomptés par le commanditaire. Il s'agissait d'une étude à propos des effets psychologiques d'un médicament « X » sur une population d'adolescents handicapés mentaux, mais ne présentant pas de perturbation neurologique décelable à l'E.E.G. et à l'examen clinique.

La stratégie de cette étude fut la suivante. Le superviseur (P1) du service psychologique de l'institution demande au psychologue (P2) d'appliquer aux adolescents une batterie de tests afin de préparer les dossiers de routine, comme il demande au psychiatre de lui indiquer les sujets présentant des perturbations neurologiques. P2 applique les tests et transmet les résultats à P1. Celui-ci, sans avoir vu les adolescents, mais à partir des données fournies par le psychiatre et des résultats obtenus par P2, constitue des paires de sujets en utilisant l'indice de similitude de Sørensen (tel que présenté par Dagnelie). Mise au courant du nombre de paires composées, la firme expédie à P1 les flacons de pilules par paires, A et B, gardant pour elle le code de répartition du médicament et du placebo. P1, ne sachant donc quel flacon de chaque paire contient le médicament ou le placebo, y étiquette les noms des adolescents selon les paires qu'il a composées, et ensuite transmet les flacons à l'éducateur chargé de la distribution des médicaments. D'autres éducateurs (stagiaires) sont priés d'observer et d'annoter une série de comportements sur une liste sans qu'ils sachent à quelle partie de la paire appartient l'adolescent qui leur est assigné. À la fin de la période d'expérimentation, P2 applique à nouveau les tests et en transmet les résultats à P1 qui reçoit des éducateurs les fiches d'observation complétées. Ensuite P1, ayant ces résultats, établit les différences absolues pour chaque paire, puis recevant seulement alors de la firme le code de distribution du médicament et du placebo dans les paires de flacon, il établit le sens des différences et peut tester la signification statistique des changements intervenus (dans ce cas par un Wilcoxon matched-pairs signed ranks test).

Quelle est la position de chacun des intervenants dans cette expérimentation ?

— P1, le superviseur du service, ne voit pas les adolescents, travaille uniquement sur des chiffres, et jusqu'au moment où il calcule le Wilcoxon, ne sait pas à quelle série de données (partie des paires) correspond le traitement. Il sait quels types d'effets sont attendus du produit, ayant discuté avec la firme des éléments de la liste d'observation et des tests psychologiques à retenir (le P.F. Rosenzweig, le Z et le style moteur).

— P2, le psychologue, rencontre et teste les adolescents, mais ne sait qui est pairé avec qui (il n'est pas au courant du principe d'établissement des paires qui a été choisi) et donc ignore qui reçoit le médicament ou le placebo. Il sait seulement qu'une expérimentation est en cours mais ignore les effets attendus du produit ; et pour qu'il ne puisse les déduire, P1 ne lui a pas dit lesquels parmi les tests de la batterie ont été retenus pour la recherche.

— Les éducateurs-observateurs, comme le responsable de la distribution des médicaments, sont dans la même situation que P2. Les items de la liste d'observation couvrant un large éventail d'activités, il leur est aussi difficile d'en déduire les effets attendus.

— Les adolescents se rendent compte qu'on leur donne à tous une pilule supplémentaire. Peut-être certains s'interrogent-ils à ce sujet et remarquent que les éducateurs sont peut-être plus attentifs à leurs comportements. Mais P2 et les éducateurs ne sachant qui est du groupe expérimental ou du groupe témoin et ne sachant quels sont les effets attendus (sinon une amélioration, mais sur quel plan ?), les indices à ce sujet sont semblables pour tous. Dès lors, si suite à ces interactions des modifications du comportement apparaissent, elles ne devraient pas être systématiquement orientées chez les uns plus que chez les autres.

— La firme pharmaceutique sait dans quels flacons est le médicament ou le placebo. Mais elle ne connaît ni la composition des paires ni les individus qui reçoivent le produit actif, et elle n'est en contact ni avec les adolescents ni avec P2 et les éducateurs.

Ainsi, pendant cette expérimentation ce n'était pas deux mais cinq groupes d'aveugles qui opéraient, chacun ignorant une partie de ce qu'il aurait pu souhaiter savoir.

En utilisant ce schéma, nous n'avons pas pu éliminer toutes les occasions de contamination envisagées plus haut. Une majeure partie d'entre elles l'ont cependant été. C'est ce que nous allons examiner avant d'envisager quel schéma idéal eut dû être mis en place.

Parmi les quatre premières possibilités de contamination (A. recueil des données), celle à laquelle nous risquons d'avoir le plus donné prise est la perception sélective (A.4.). Les items de la grille d'observation et les tests psychologiques avaient été choisis en regard de certaines hypothèses : ils étaient censés pouvoir dépister les effets principaux du médicament. Les résultats nous ont conduit à conclure que, dans les conditions expérimentales données, aucun effet n'était apparu. Mais en fait ceci ne vaut que pour les indices, les « comportements-signes » que nous avons choisis. D'autres tests, d'autres items n'auraient-ils pas révélé certains effets... secondaires ? Secondaires pour nous à ce moment, mais peut-être importants par ailleurs.

Quant aux trois autres possibilités (A.1. à A.3.), les effets en sont neutralisés, car étant donné la méthode utilisée, si elles s'appliquent, c'est à tous les éléments de la population, témoins aussi bien qu'expérimentaux. Par ailleurs, si dans ce cadre elles ont pu jouer, c'est plus en fonction de rapports inter-individuels (adolescent-testeur ou observateur) qu'en fonction de rapports individu-(observateur ou testeur) groupes (témoin ou expérimental).

Si l'on examine les possibilités de contamination des résultats (B), nous pouvons poser que la méthode des aveugles a permis d'échapper aux contaminations relevant des réactions aux données de la situation, que ce soit chez les sujets ou chez les expérimentateurs. De plus, l'effet Rosenthal (B.3.) y est neutralisé, s'il n'est pas absent : le manque absolu de contact entre le chercheur principal et les sujets minimise l'influence des hypothèses sur les résultats. Si cependant cette influence peut « passer » par l'intermédiaire des observateurs ou du testeur, elle se voit neutralisée, puisque ces derniers ne savent pas à quel groupe le sujet qu'ils ont devant eux appartient. Ils ne peuvent ni lui coller une étiquette, ni s'inspirer de son histoire expérimentale, pour aligner leurs modes d'interaction avec lui.

Cependant, telle qu'utilisée, la méthode des aveugles ne permet pas d'échapper aux contaminations réactives au fait expérimental (B.1.1. et B.2.1.). Il n'est guère possible d'introduire l'usage d'une nouvelle grille et la distribution généralisée d'une pilule sans que les éducateurs-observateurs et les adolescents-sujets ne se rendent compte qu'il « se passe quelque chose ». Dans un établissement semblable à celui où s'est déroulée cette expérimentation, les examens neurologiques et psychologiques sont monnaie courante. Pour ces adolescents, dont certains avaient pratiquement toujours vécu en institution, le test est une routine à laquelle bien souvent ils se soumettent sans poser la question du pourquoi, du rythme, etc... Mais si un beau jour ils constatent que tous reçoivent en même temps et aux mêmes heures « une pilule » en plus, on peut s'attendre à une sensibilisation et à une modification des comportements. Quant aux éducateurs, s'étant vu demander de remplir une nouvelle grille, applicable à tous les sujets et exigeant une attention particulière de leur part, ils auront eu leur routine perturbée. En conséquence, ils auront sans doute manifesté et développé de nouveaux types d'interactions avec les adolescents, interactions dont le style aura probablement été limité à la durée de l'expérience. Bien sûr, tant chez les adolescents-sujets que chez les éducateurs-observateurs, ces contaminations ont été effectives pour les deux groupes, expérimental et témoin. Ceci ne favorise pas la confirmation des différences entre ces groupes et permet donc de moins facilement soutenir l'hypothèse d'un effet du traitement. Du moins dans ce cas, peut-on penser que le résultat observé est fonction non seulement du traitement expérimenté (le médicament), mais aussi du traitement secondaire dû à la situation expérimentale (modification des interactions éducateurs-adolescents). À la limite, on pourrait se demander si le traitement secondaire n'a pas eu des effets plus puissants que le traitement expérimental. Cela expliquerait l'absence de différence

observée : la modification des interactions ayant un effet tel sur les comportements que la modification médicamenteuse supplémentaire ne peut être qu'imperceptible.

Pour échapper à cette contamination, qui par ailleurs conduit dans ce cas à la prudence quant à l'usage du médicament et donc pourrait être considérée comme bénéfique, il eut sans doute fallu paier les sujets non pas deux par deux mais trois par trois, ou plus encore. Mais le trois par trois nous eut suffi si nous avions pensé à paier, sur base des tests de routine, un troisième sujet pris dans une autre section de l'institution où traitement, pseudo-traitement et observation continue n'auraient pas été effectués, nous contentant pour ce groupe du post-test. Nous perdions alors l'information issue des grilles et nous devons alors surmonter la difficulté de rendre « aveugle » le testeur quant à l'appartenance des sujets à ce troisième groupe. Cependant, si cette difficulté avait pu être surmontée, nous aurions bénéficié d'une observation terminale au niveau des tests psychologiques qui nous aurait permis de mesurer l'effet du traitement secondaire (différences entre les résultats du groupe à pseudo-traitement — le placebo — et du groupe sans traitement).

Enfin, en ce qui concerne les autres sources de contamination (B.4. et C.), la méthode des aveugles les a neutralisées dans la mesure où leurs composantes ont été contrôlées. Or nous venons de voir ce qu'il eut fallu faire pour mesurer l'effet des contaminations B.1.1. et B.2.1.

Dans des recherches du type envisagé (étude de l'effet d'un traitement, qu'il soit pharmaceutique, psychothérapeutique ou pédagogique), un schéma idéal devrait s'inspirer du plan à 4 groupes de Solomon (voir D.T. Campbell). Il conviendrait alors de procéder en utilisant six (6) groupes considérés comme équivalents de par leurs caractéristiques générales (à partir d'épreuves de routine, de catégories administratives, cliniques, socio-économiques ou autres). De ces six groupes, une première série de trois subirait pré-test, appariement et post-test, les trois autres ne subissant que le post-test (ceci pour contrôler la réactivité au fait expérimental). Mais dans les deux séries un même schéma subsisterait quant au traitement : n° 1 et 4 avec le traitement à expérimenter, n° 2 et 5 avec un pseudo-traitement, n° 3 et 6 sans traitement. Par ailleurs, ceci devrait se faire en mettant en situation d'« aveugle » l'observateur qui prend les mesures (initiales, en cours et terminales) ainsi que les expérimentateurs (éducateurs, enseignants, etc...) qui appliquent les traitements et les pseudo-traitements.

Un tel schéma peut se représenter de la manière suivante.

Ce schéma peut se lire de la manière suivante. À partir d'une population déterminée, le chercheur principal (celui qui veut vérifier l'une ou l'autre hypothèses) effectue une opération d'allocation des sujets en six groupes de manière à ce que ceux-ci soient équivalents, et cela à partir de données recueillies antérieurement, mais pertinentes par rapport aux dimensions impliquées par les hypothèses. Un observateur, ignorant la

|                     | groupe | observ.<br>initia. | appar. | trait.<br>avec<br>ou ss.<br>observ. | observ.<br>termin. |                   |
|---------------------|--------|--------------------|--------|-------------------------------------|--------------------|-------------------|
| popul. — allocation | n° 1   | 0.1                | r. 1   | X                                   | 0.2                | comparai-<br>sons |
|                     | n° 2   | 0.3                | r. 2   | x'                                  | 0.4                |                   |
|                     | n° 3   | 0.5                | r. 3   | —                                   | 0.6                |                   |
|                     | n° 4   | —                  | —      | X                                   | 0.7                |                   |
|                     | n° 5   | —                  | —      | x'                                  | 0.8                |                   |
|                     | n° 6   | —                  | —      | —                                   | 0.9                |                   |

symboles : 0. (1 à 9) : observations ou tests  
 r. (1 à 3) : appariements  
 X : traitement  
 x' : pseudo-traitement  
 — : absence d'intervention

technique et/ou la règle d'allocation des sujets dans les groupes et les hypothèses mises en question, effectue les observations initiales et, quand il y a lieu, les observations en cours d'expérimentation pour les groupes 1, 2 et 3. Maintenu dans la même ignorance il effectuera les observations finales pour l'ensemble des groupes. Ayant communiqué les résultats de l'observation initiale au chercheur principal, celui-ci pourra établir l'appariement des sujets des groupes 1, 2 et 3. Les traitements (X, x'), donnés aux sujets des groupes 1, 2, 4 et 5, le sont sans que les agents qui effectuent ces traitements (éducateurs, enseignants ou autres) sachent si ce qu'ils font correspond au traitement (X ou action dont on ne connaît pas les effets mais dont on les suppose et qui sont à vérifier) ou au pseudo-traitement (x', action sans effet ou dont on connaît les effets, ceux-ci n'étant pas l'objet des hypothèses de la recherche). De plus, ces agents ne peuvent savoir à quel groupe appartiennent les sujets auprès de qui ils opèrent, ni quels types de groupe sont inclus dans la recherche. Après le traitement et l'observation terminale, le chercheur ayant effectué au début les opérations d'allocation et d'appariement, peut procéder aux comparaisons des résultats après avoir reçu les résultats de l'ensemble des observations. Il peut effectuer ces comparaisons en deux temps. D'une part, il compare les gains réalisés ou non par les sujets appariés dans les trois premiers groupes (soit 0.1 — 0.2 / 0.3 — 0.4 / 0.5 — 0.6). Cela lui permet d'isoler l'effet du traitement (X) de l'effet possible d'une contamination par perception des consignes implicites quant au statut des sujets (X et —) et quant aux hypothèses (X et x'). D'autre part, il confronte les résultats obtenus, après la période de traitement, par chacun des six groupes, les comparaisons de 0.2 avec 0.7, de 0.4 avec 0.8 et de 0.6 avec 0.9 lui

permettant de déterminer l'importance de la réactivité des agents de traitement comme des sujets au fait expérimental. Étant entendu que ces comparaisons s'effectuent sans que le chercheur ait pu avoir une influence sur l'interaction observateur-sujets et sur l'interaction agent de traitement-sujets.

Notons en outre que si les données des observations doivent être codées ou interprétées avant de pouvoir être utilisées par le chercheur, il conviendrait, pour ce faire, de recourir à un codeur ou à un interprète n'ayant pas vécu l'observation elle-même et ignorant les hypothèses ainsi que la règle d'allocation des sujets dans les groupes. Cette mesure devrait permettre de contrôler les effets de stéréotypie et les effets de halo, surtout en ce qui concerne les interférences entre les faits d'observation vécus mais non notés (par exemple, faits non répertoriés sur la grille) et les faits d'observation notés. Encore pourrait-il être utile de contrôler les effets de familiarisation avec le système de codage en faisant recoder certaines données de l'observation initiale après le codage des données de l'observation finale (fidélité intra-codage). Ceci n'élimine pas la nécessité d'établir la fidélité des observations par la comparaison de la notation des faits par différents observateurs (fidélité inter-observateur) et, lorsque le matériel a pu être enregistré (au magnéscope par exemple), par la comparaison des notations successives, mais espacées dans le temps, d'une même séquence par le même observateur (fidélité intra-observateur). Soulignons enfin qu'en ce qui concerne les observations ou les tests initiaux et terminaux, il serait utile de recourir à des observateurs ou testeurs différents pour chacun des stades afin d'éviter des contaminations par « souvenir ». Le fait d'avoir vu l'enfant une première fois laisse une empreinte qui interfère avec l'observation ultérieure, et souvent la biaise.

Un tel schéma n'est sans doute pas encore parfait, mais surtout il n'est pas toujours praticable. Un premier problème se pose : l'effectif de la population de base à partir de laquelle les différents groupes doivent être extraits : plus le nombre de groupes augmente, plus l'effectif-source doit être grand. Dès lors, dans bien des cas, le chercheur devra sacrifier l'un ou l'autre des groupes parmi ceux de la seconde série. En outre, cela multiplie les observateurs-testeurs de même que les agents qui appliquent le traitement et le pseudo-traitement. Or, plus on recourt à un grand nombre de collaborateurs, plus on risque de « vendre la mèche » et de transformer les « aveugles » en « semi-voyants ». Enfin, en situation scolaire, cela implique une multiplication des locaux et des agents pédagogiques (administrateurs et enseignants), ce qui rend plus difficile le contrôle des facteurs de l'environnement humain et physique.

Les imperfections qu'un tel schéma peut encore comporter (et que plus d'un prendra plaisir à souligner, et, il faut l'espérer, à corriger), ainsi que ses difficultés d'application, n'en font qu'un modèle dont on peut s'inspirer et non pas un modèle idéal. D'ailleurs, si le modèle idéal existait, la méthodologie de la recherche cesserait d'être une question de débat, d'enseignement et de préoccupation.

## RÉFÉRENCES UTILES

- Dagnelie P., 1968, Introduction aux problèmes et aux méthodes de classification numérique. *Biométrie-Praximétrie*, IX, 2, 87-111.
- De Landsheere G., 1970, *Introduction à la recherche en éducation*. Liège, G. Thone. II<sup>e</sup> partie : l'observation et la mesure des comportements, 39-54.
- Delay J., Perse J., Pichot P., 1966, *Méthodes psychométriques en clinique*. Masson, Paris. Livre II : La validité des tests de personnalité en psychiatrie. 93-114.
- Lemaine G., Lemaine J.M., 1969, *Psychologie sociale et expérimentation*. Paris, Mouton-Bordas :
- Matalon B., La logique des plans d'expérience, 35-45.
  - Campbell D.T., Facteurs intéressant la validité des études expérimentales dans des contextes sociaux, 47-61.
  - Rosenthal R., La participation volontaire, 71-79.
  - Riecken, H.W., Programme de recherches sur l'expérimentation en psychologie sociale, 271-281.
  - Orne M.T., Psychologie sociale de l'expérimentation en psychologie : les consignes implicites et leurs conséquences, 283-290.
  - Rosenthal R., L'influence de l'expérimentateur sur les résultats de la recherche en psychologie, 291-310.
- Rosenthal R., Rosnow R.L., 1969, *Artifact in behavioral research*. New York, Academic Press.
- Rosenthal R., Jacobson, L., 1971, *Pygmalion à l'école*, Tournai, Casterman.