



LA META-ANALYSE DES RECHERCHES EXPERIMENTALES EN EDUCATION

par D. LECLERCQ

A. Pourquoi résumer les analyses?

Au cours des deux derniers siècles, le volume des publications scientifiques a doublé tous les quinze ans. Les efforts d'intégration, des synthèses permettant des vues d'ensemble, sont évidemment devenues nécessaires. Prenons comme exemple un thème particulier : le rendement des systèmes d'individualisation de l'apprentissage. La tradition scientifique anglosaxonne offre quantité de recherches méthodologiquement bien menées, présentant des résultats soigneusement contrôlés (groupes témoins de capacités équivalentes au départ) portant sur de nombreuses variables dépendantes et indépendantes. Dans un tel contexte, il devient envisageable de fournir une synthèse statistique de centaines de recherches venant d'horizons différents. C'est le principe de la méta-analyse.

Alors qu'une dizaine d'études peuvent suffire pour résoudre un problème en biologie, ce nombre est souvent largement insuffisant en éducation. Les résultats dépendent en effet d'une gamme plus importante de facteurs... que l'on sait moins bien repérer et contrôler. La prolifération des recherches est, pour quelque temps encore, inévitable.

Le raisonnement de GLASS (1976, p. 4) est le suivant : « Avant que les résultats puissent être utilisés, qu'ils puissent persuader les sceptiques, influencer les politiques, affecter les pratiques, ils doivent être connus. Quelqu'un doit les organiser, les intégrer, en extraire le message. Une centaine de doctorats sont muets. Quelqu'un doit les lire et découvrir ce qu'ils disent ».

GLASS (1976) cite aussi une phrase du poète Thomas Stern ELLIOT : « Où est la connaissance que nous avons perdue par l'information ? »

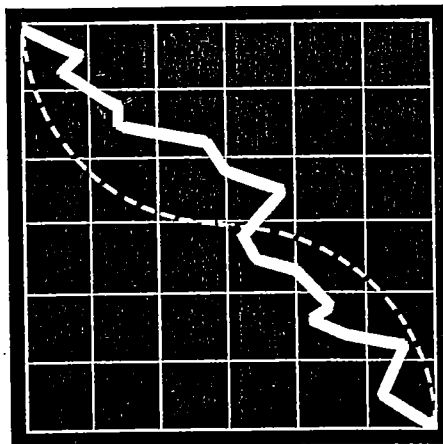
Par là, GLASS veut souligner un paradoxe, celui de « l'explosion de la mésinformation »⁽¹⁾. Nous sommes inondés d'informations⁽²⁾. Notre

(1) « Mésinformation explosion ».

(2) Chaque année, le système ERIC répond à deux millions de demandes.

Dieudonné LECLERCQ

Le présent document est organisé autour d'une synthèse de divers travaux menés par J. KULIK et ses collègues du C.R.L.T. dont le LOGO ci-dessous oppose la progression irrégulière de l'individu (ligne brisée) à la courbe lisse de la progression collective (statistique) vers la maîtrise.



*The Center for Research
on Learning and Teaching*

THE UNIVERSITY OF MICHIGAN
ANN ARBOR

problème est d'y pêcher la connaissance. Nombreux sont les auteurs qui résolvent ce problème en éliminant les recherches qui ne présentent pas les quelques caractéristiques auxquelles ils sont attachés. Il ne reste souvent alors que *leur* recherche et celles de quelques-unes de leurs meilleures relations professionnelles.

Pour GLASS, la différence de validité des résultats entre des expériences « bien » ou « mal » conçues est faible et éliminer les secondes revient à se priver d'une quantité considérable de données importantes. Ce manque de rigueur apparent dans la sélection des expériences ne se retrouvera pas chez des chercheurs comme KULIK et ses collaborateurs (Université de Michigan à Ann Arbor).

B. Analyses primaires et secondaires

L'analyse primaire est ce que chacun imagine de l'analyse statistique des données d'une recherche.

L'analyse secondaire consiste à ré-analyser des données soit pour répondre à de nouvelles questions, soit pour appliquer d'autres traitements (en principe meilleurs) aux mêmes données de base.

L'analyse secondaire a ses lettres de noblesse :

- MOSTELLER et MOYNIHAN (1972) réanalysant l'étude de COLEMAN *et al.* sur l'égalité des chances devant le système éducatif américain, en relation avec des facteurs économiques et raciaux.
 - CAMPBELL et ERLEBACHER (1970) réanalysant l'évaluation faite par le groupe OHIO-WESTINGHOUSE du projet HEADSTART portant sur l'éducation compensatoire précoce en faveur des enfants « désavantagés ».
 - ELASHOFF et SNOW (1971) réanalysant les données de ROSENTHAL et JACOBSON concernant l'effet PYGMALION, c'est-à-dire l'auto-réalisation des prédictions.
- On sait que des organismes, par exemple l'IEA ⁽³⁾ ou le NAEP ⁽⁴⁾ prennent des précautions particulières (archives soignées des données sur bandes magnétiques) pour faciliter les analyses secondaires, qu'ils encouragent même au moyen de bourses.
- Pour GLASS, il est exclu de fonder la méta-analyse sur un traitement (secondaire) des données de base des études, car ces données sont dans de nombreux cas inaccessibles.

⁽³⁾ Association for the International Evaluation of Educational Achievement.

⁽⁴⁾ National Assessment of Educational Progress.

C. Les principes de la méta-analyse

1. Définition :

G.V. GLASS (1976) a, le premier, employé ce terme pour désigner l'analyse systématique d'analyses, ou plus précisément, « l'analyse statistique d'une grande quantité de données tirées de différentes études pour en intégrer les résultats » (p. 3).

2. Détermination de l'Ampleur de l'Effet ⁽⁵⁾

D'une recherche à une autre, les résultats ne sont pas comparables : les tests employés sont différents (donc les maxima possibles, les moyennes, les dispersions et les différences entre groupes).

Quand les données des études le permettent, il est possible de surmonter cet écueil (notamment dans la recherche de BANGERT *et al.*, 1983 décrite ci-après).

N. GLASS (1976) a proposé de « standardiser » les résultats en calculant l'ampleur de l'effet (A.E.) par la différence entre la moyenne du groupe expérimental (Me) et la moyenne du groupe témoin (Mt), différence divisée par l'écart-type du groupe-témoin (σ) :

$$A.E. = \frac{Me - Mt}{\sigma}$$

Si les rapports ne fournissent pas les ingrédients de base (Me, Mt et σ), l'E.S. est calculé à partir d'indices comme le t ou le F, par une technique exposée par MCGRAW et GLASS (1980), sinon par des techniques de régression. En plus de cette « ampleur » de l'effet, son erreur standard et le seuil de signification de cette ampleur de différence sont calculés.

Cette technique de présentation des résultats nous paraît particulièrement contribuer au courant « éducatif », car on y accorde autant d'attention à l'importance pédagogique d'une différence (en vaut-elle la peine ?) qu'au degré de signification (est-elle due au hasard ?).

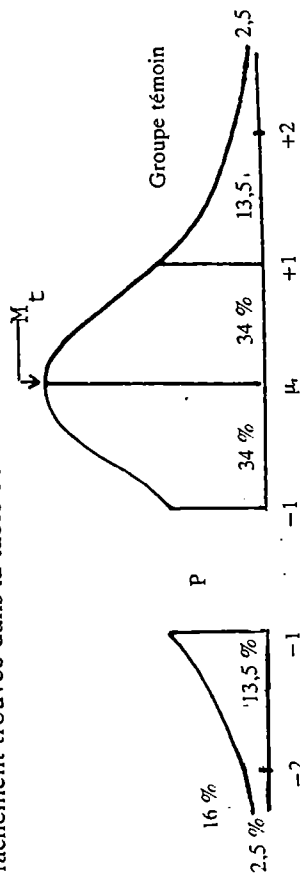
En outre, le fait que l'A.E. soit exprimé en unités d'écart-type, facilite la représentation graphique et l'interprétation de cet indice. On peut en effet illustrer, de manière purement théorique, l'A.E. observé en dessinant deux courbes de Gauss (aux portions d'aires bien connues) dont les moyennes sont distantes d'une A.E. (d'une ampleur d'effet) donnée.

⁽⁵⁾ « Effect size » (ou E.S.) en anglais.

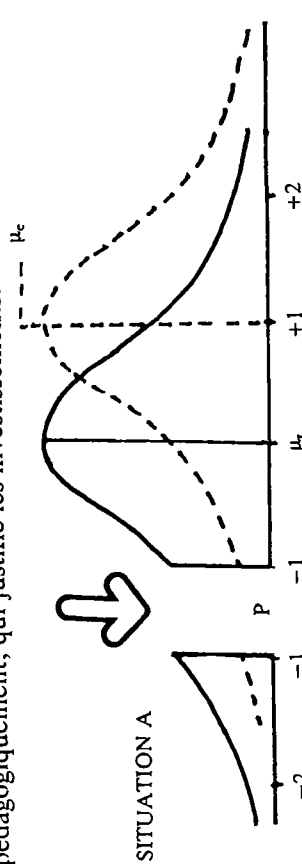
« Ampleur de l'effet » (ou A.E.) en français.

3. Interprétation de l'ampleur de l'effet (A.E.)

Plaçons pour une matière donnée, le « passing score » (P) à la moyenne (du groupe témoin) moins un écart-type (σ). Cela signifie que, 16 % des élèves de ce groupe-témoin ne réussissent pas. Cette valeur peut être facilement trouvée dans la table de la loi normale.



Considérons une expérience A au terme de laquelle le groupe expérimental obtient des résultats supérieurs à ce groupe témoin : l'ampleur de l'effet est d'un écart-type. Le score de passage (indiqué par la lettre P) se trouve à un écart-type plus à gauche pour le groupe expérimental (traits interrompus). Donc, seulement 3,5 % des élèves (au lieu de 16 %) ne réussissent pas. Avouons que c'est là un résultat qui « en vaut la peine » pédagogiquement, qui justifie les investissements.



Considérons une expérience B où l'ampleur de l'effet en faveur du groupe expérimental n'est que de 0,5 écart-type. Seulement 9,5 % des étudiants ne réussiront pas (au lieu de 16 % dans le groupe contrôle).

Considérons enfin une expérience C où l'ampleur de l'effet en faveur du groupe expérimental n'est que de 0,1 écart-type; 14,5 % des étudiants

(*) Le « passing score » est le score minimal de réussite, celui en dessous duquel l'étudiant redouble la classe, par exemple.

échoueront (contre 16 % dans le groupe témoin) et on peut se demander si l'énergie *additionnelle* consentie pour l'expérience n'aurait pas pu être *mieux investie* !

GLASS, McGAW et SMITH (1981, p. 103) ont proposé de discuter en termes d'*unités exprimées en « grades »* (aux U.S.A., ce terme désigne les années scolaires). Leur raisonnement est que, par définition, l'élève moyen « gagne » dix mois de compétence durant une année scolaire. Or, par ailleurs, on a constaté que dans la plupart des tests de rendement à l'école primaire, l'élève d'une classe (par exemple la 4^e primaire) a un score supérieur d'un *écart-type* par rapport à l'élève moyen de la classe inférieure (par exemple la 3^e primaire). Par conséquent, l'ampleur de l'effet (A.E.) d'une année d'enseignement au niveau primaire est grosso modo de + 1, et donc celle d'un mois est, en moyenne, de + 0,1. Cela revient à ce que l'individu moyen passe du 50^e au 54^e percentile (voir en annexe une table de conversion entre les percentiles et les ampleurs d'effet).

D. Les étapes d'une méta-analyse

Voici les diverses étapes d'une telle démarche. Les points 1, 2 et 3 sont illustrés par les étapes de l'étude de BANGERT *et al.* (1983) sur les effets de l'enseignement individualisé à l'école secondaire.

1. Interrogation de banques de données

Les auteurs commencent par une *interrogation bibliographique*.

Par exemple, BANGERT *et al.* ont consulté les banques de données gérées par LOCKEED.

a) ERIC : Educational Resources Information Center

i — Research in Education

ii — Current Index to Journals in Education

b) Les Psychological Abstracts

Les auteurs ont trouvé 450 articles sur l'enseignement individualisé, mais seulement quelques-uns semblaient donner les résultats de recherches expérimentales.

2. Obtention des articles qui semblent les plus intéressants

Certains figurent dans la bibliographie des articles de départ (c'est le principe de la bibliographie-boule-de-neige).

On peut regretter l'absence d'un questionnaire détaillé aux enseignants leur demandant de fournir, pour chaque objectif, l'importance qu'ils lui attribuent et la manière dont il a été enseigné ou appris (*opportunity to learn*). Cette lacune est évidemment due aux études de base, et non aux chercheurs qui procèdent à la méta-analyse.

4. Détermination des variables dépendantes

Il est rare que l'on dispose de recherches fournissant des résultats sur plus de trois ou quatre variables dépendantes.

Voici certaines variables recueillies dans diverses études (BANGERT et al. 1983) KULIK et al. 1979. a et 1979. b).

- a) Le rendement des élèves, mesuré par des examens sur la matière scolaire
 - i — *immédial*, c'est-à-dire une ou deux semaines après la fin du programme (49 des 51 études de BANGERT et al.).
 - ii — à *moyen terme*, c'est-à-dire plusieurs mois après la fin du programme : mesure la *rétenion* (8 des 75 études de KULIK et al., 1979. a).
 - iii — par recours à *des questions fermées ou courtes*, par exemple à des questions à choix multiple ou à *des questions de production longue*, à correction complexe.
- b) *La pensée critique* mesurée par des tests *ad hoc*. Il s'agit là de ne pas se limiter à des effets « convergents » (4 des 51 études de BANGERT et al.).
On peut regretter de ne pas trouver ici d'étude des effets de l'enseignement individualisé sur les *méthodes de travail*.
- c) *L'attitude des élèves vis-à-vis de la matière* mesurée sur des échelles *ad hoc* (4 des 51 études de BANGERT et al. ; 11 des 75 études de KULIK et al., 1979 a).
On peut regretter de ne pas trouver, parallèlement à cette variable, une mesure de « satisfaction » de l'enseignant sur la méthode.
- d) *L'image de soi* des élèves (4 des 51 études de BANGERT et al.).
On peut regretter de ne pas trouver ici des mesures d'autoestimation (réalisme) et d'autorégulation (autonomie).
- e) *La persévération* (course completion) mesurée par le pourcentage d'élèves qui vont jusqu'au bout du cours, par rapport au nombre d'inscrits (27 sur 75 des études de KULIK et al., 1979 a).
- f) *La durée*, mesurée par le temps moyen consacré par les élèves à l'apprentissage (8 des 54 études de KULIK et al., 1979 c).

BANGERT et al. ont ainsi obtenu 51 études répondant à leurs critères :

- a) Un système d'enseignement individualisé devait être mis en œuvre dans une (ou des) classe(s) expérimentale(s) tandis que le système traditionnel était en vigueur dans la (ou les) classe(s) contrôle(s).
- b) Ces études devaient être faites dans de vraies classes du secondaire (ce qui exclut les études en laboratoire).

c) Ces études devaient rendre compte de résultats mesurés quantitativement dans les deux groupes, expérimental et témoin.

d) Ces études devaient être exemptes de faiblesses méthodologiques comme, par exemple,

- i — une forte différence d'aptitude entre les deux groupes comparés
 - ii — une préparation au test offerte injustement à l'un des groupes
- e) Ces études devaient avoir été publiées après 1955.

3. Description des variables indépendantes

Pour vérifier les facteurs qui pourraient influencer les résultats, BANGERT et al. ont arbitrairement décidé de « caractériser » les études selon 12 variables (chacune étant dichotomique ou trichotomique), ici :

1. Le rythme : 1. de groupe 2. individuel
2. Maîtrise exigée ? 1. non 2. oui
3. Groupement selon test ? 1. non 2. oui
4. Avec vidéo ? 1. non 2. oui
5. Equivalence des participants 1. sans preuve 2. oui 3. hasard
6. L'enseignant dans les deux groupes est 1. différent 2. le même
7. Durée de l'expérience : 1. moins de 19 semaines 2. 19 semaines et plus
8. Niveau scolaire : 1. secondaire inférieur 2. secondaire supérieur
9. Nombre d'étudiants : 1. 15 et moins 2. 16 et plus
10. Matière : 1. Mathématique 2. Sciences 3. autre
11. Source de la description : 1. non publié 2. doctorat 3. publié
12. Année de publication : 1. entre 1955 et 1970 2. 1971 et après

Certaines de ces caractéristiques (par exemple : 2, 5, 6, 7, 8, 12) étaient définies *avant* l'examen des documents, d'autres (par exemple : 4, 9, 10...) *après*. Ainsi, les auteurs ont trouvé 28 études sur les mathématiques, 16 sur les sciences et 5 « autres » seulement.

Cette variable pourrait être considérée comme «indépendante». Néanmoins, dans une perspective de la maîtrise, il s'agit souvent du «revers» du rendement (voir détails dans LECLERCQ, DONNAY, DE BAL, 1977, p. 168 et p. 194).

g) *La corrélation rendement-aptitudes*

De nombreuses études ont montré que les étudiants présentant des meilleures aptitudes au départ bénéficient plus que les autres d'un cours collectif (Y. BEGIN, 1981). On devrait s'attendre (selon les hypothèses de BLOOM, 1981, par exemple), à ce qu'un système individualisé basé sur la maîtrise diminue la corrélation (habituellement supérieure à 0,5) entre le rendement et les aptitudes. *On peut regretter que jamais ne soit étudiée l'interaction aptitude-traitement* (recherche A.T.I.) quand l'aptitude est définie autrement qu'un score (quantitatif) à un test.

Dans un prochain article intitulé «Le rendement des Utilisations pédagogiques de l'ordinateur», l'intérêt de la méta-analyse sera illustré, de même que ses limites.

D. LECLERCQ

Chercheur au Laboratoire de Pédagogie expérimentale - ULg

Bibliographie

- COOK, T.D., The potential and limitations of secondary evaluations. Chapter 6, pp. 155-234. in APPLE, M.W. SUBKOVIAK, H.S., and LUFLE, J.R. (Eds), *Educational Evaluation: Analysis and responsibility*. Berkeley: Mc Cutchan, 1974.
- CAMPBELL, D.T. and ERLEBACHER, A.E., How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In *Compensatory education: A national debate*. Hellmuth, J. (Ed.) The Disadvantaged Child., Vol. 3. New York: Brunner/Mazel, 1970.
- ELASHOOF, J.D., and SNOW, R.E. (Eds.). *Pygmalion reconsidered*. Worthington. Ohio: Charles A. Jones, 1971.
- GLASS, G., *Primary, secondary, and meta-analysis of research*, Educational Researcher, 1976, vol. 5, p. 3-8.
- GLASS, G.V., Mc GAW, B & SMITH, M.L. (1981). *Meta-analysis in social research*. Beverly Hills, C.A.: Sage Publications.
- JACKSON, G.B. The research evidence on the effects of grade retention. *Review of Educational Research*, 1975, 45, 613-635.
- KULIK, J.A., BANGERT, R.L. & WILLIAMS, G.W., (1983). Effects of computer-based teaching on secondary school students. *Journal of Educational Psychology*, 75, 19-26.
- LIGHT, R.J., and SMITH, P.V. Accumulating evidence: Procedures for resolving contradictions among different research studies. *Harvard Educational Review*, 1971, 41, 429-471.
- MOSTLELLER, F.M. and MOYNIHAN, D.P. (Eds), *On equality of educational opportunity*. New-York: Vintage Books, 1972.
- SCHRAMM, W., Learning from instructional television. *Review of Educational Research*, 1962, 32, 156-167.