

L'échantillonnage et le problème de la validité externe de la recherche en éducation

Djavid Ajar, Clément Dassa and Hélène Gougeon

Volume 9, Number 1, 1983

URI: <https://id.erudit.org/iderudit/900396ar>

DOI: <https://doi.org/10.7202/900396ar>

[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

Ajar, D., Dassa, C. & Gougeon, H. (1983). L'échantillonnage et le problème de la validité externe de la recherche en éducation. *Revue des sciences de l'éducation*, 9(1), 3–21. <https://doi.org/10.7202/900396ar>

Article abstract

This paper discusses the relationship between sampling theory and external validity in research in the human sciences, particularly in education. The conclusions reported in doctoral and master's theses in the fields of psychology and education in two Quebec universities are examined. A study of the nature and quality of generalization of these conclusions is used to illustrate the practical problems regarding sampling. The size and importance of the problems identified is used as a background for some remedial suggestions.

L'échantillonnage et le problème de la validité externe de la recherche en éducation

Djavid Ajar, Clément Dassa et Hélène Gougeon*

Résumé — Les liens existant entre la théorie de l'échantillonnage et la validité externe des recherches en sciences humaines, et plus particulièrement en éducation, sont exposés. Une étude sur la nature et la qualité de la généralisation des conclusions présentées dans des thèses de doctorat et des mémoires de maîtrise en psychologie et en sciences de l'éducation de deux universités québécoises sert à illustrer les problèmes d'échantillonnage rencontrés en pratique. L'ampleur et l'importance des lacunes mises en évidence servent de fond à quelques suggestions de correctifs.

Abstract — This paper discusses the relationship between sampling theory and external validity in research in the human sciences, particularly in education. The conclusions reported in doctoral and master's theses in the fields of psychology and education in two Quebec universities are examined. A study of the nature and quality of generalization of these conclusions is used to illustrate the practical problems regarding sampling. The size and importance of the problems identified is used as a background for some remedial suggestions.

Resumen — El presente artículo expone los nexos existentes entre la teoría de muestreo y la validez externa de las investigaciones en las ciencias humanas, muy particularmente en educación. Un estudio sobre la naturaleza y la calidad de la generalización de las conclusiones obtenidas en tesis de doctorado y de maestría en psicología y ciencias de la educación de dos universidades de Quebec, sirven para ilustrar los problemas de muestreo encontrados en la práctica. La extensión y la importancia de las lagunas puestas en evidencia sirven de tela de fondo para sugerir ciertas correcciones.

Zusammenfassung — Dargestellt werden die Beziehungen zwischen der Theorie der Stichprobenerhebung und der äusseren Validität der Forschungen auf dem Gebiet der Humanwissenschaften, insbesondere in der Pädagogik. Eine Untersuchung der Art und der Qualität der Verallgemeinerung der Ergebnisse der Doktor- und Magisterarbeiten in Psychologie und Pädagogik an zwei Québecker Universitäten dient zur Veranschaulichung der bei der Stichprobenauswahl in der Praxis auftretenden Probleme. Die Grösse und Bedeutung der aufgezeigten Lücken ergibt die Ausgangspunkte für einige Verbesserungsvorschläge.

Introduction

L'étude des caractéristiques de la totalité d'une population est une entreprise lente et coûteuse qui, en général, fournit des informations d'une ampleur et d'une précision assez limitées¹. On évite ces difficultés en procédant à une cueillette

* Ajar, Djavid : professeur, Université de Montréal
Dassa, Clément : professeur, Université de Montréal
Gougeon, Hélène : étudiante diplômée

d'information auprès d'un sous-ensemble de la population qu'on espère représentatif. Les avantages d'une telle réduction relèvent de la *validité interne*² de la recherche et, une fois cette validité acquise, ses inconvénients relèvent de la *validité externe*².

La validité interne d'une recherche dépend essentiellement de la précision de la description des caractéristiques étudiées. En général, on peut affirmer que la validité interne d'une étude réalisée à l'aide d'échantillons est supérieure à celle d'une étude qui engloberait la population tout entière. En effet, la relative rapidité de la cueillette d'information auprès d'échantillons et l'économie qui en découle permettent d'obtenir un degré d'exactitude sensiblement supérieur à une prise d'information auprès de l'ensemble de la population (Cochran 1977, Krathwohl 1982). La qualité intrinsèque de l'information obtenue dans toute recherche est essentielle, au point qu'une validité interne acceptable constitue une condition nécessaire (mais non suffisante) pour que l'étude en question possède la qualité de la validité externe.

La validité externe est en fait la généralisabilité des résultats et est, entre autres, fonction de la représentativité des échantillons choisis. Ainsi, les études basées sur des populations entières (les recensements) qui possèdent une validité interne acceptable ont automatiquement une grande validité externe permettant de décrire les caractéristiques étudiées dans la population, tant et aussi longtemps qu'elles demeurent stables et que les conditions expérimentales restent inchangées.

Alors qu'un échantillon est une partie de la population choisie spécialement pour la représenter, l'*échantillonnage* est le processus par lequel ce choix est opéré. Il est donc évident que la généralisabilité des résultats obtenus dans toute recherche est tributaire des techniques d'échantillonnage. À ce sujet, notons que les problèmes d'échantillonnage sont habituellement plus marqués dans le domaine des sciences humaines à cause de la grande variabilité échantillonnale inhérente aux unités expérimentales reliées aux caractéristiques du comportement humain ainsi qu'à leurs instabilités spatiale et temporelle. Le problème est d'actualité comme le démontre une étude récente (Gougeon, 1982) sur les qualités de l'échantillonnage dans les recherches doctorales et de maîtrise en sciences de l'éducation et en psychologie de deux universités québécoises. Cette étude identifie les lacunes majeures de ces recherches et montre que leur validité externe laisse grandement à désirer.

Le présent article a pour but essentiel de sensibiliser les étudiants en rédaction de mémoire ou de thèse ainsi que leurs directeurs de recherche aux problèmes fondamentaux reliés à l'application inadéquate des méthodes d'échantillonnage dans le domaine de la recherche en sciences de l'éducation. C'est pourquoi nous avons jugé utile de présenter brièvement, dans un langage accessible à tous, les concepts de base de la théorie de l'échantillonnage. Un exposé théorique et technique de ces concepts pourrait faire l'objet d'un prochain article.

Dans une première partie on abordera les problèmes de la représentativité d'un échantillon, des méthodes d'échantillonnage et, finalement, de l'estimation de la taille d'un échantillon. Dans une deuxième partie, les résultats de l'étude de Gougeon (1982) serviront, à la lumière des concepts exposés dans la première partie, à illustrer les dimensions du problème de l'échantillonnage dans une importante catégorie de recherches en psychologie et en sciences de l'éducation.

I- La représentativité d'un échantillon de sujets

Dans le langage courant, la représentativité est considérée comme une propriété intrinsèque de l'échantillon. Ainsi, Larousse définit l'échantillon comme une « petite quantité d'un produit quelconque, qui sert à apprécier la qualité du tout (...) » (Grand Larousse, Tome 4). Malheureusement, en sciences humaines, la complexité et la taille des populations que l'on veut décrire sont autant d'obstacles à la représentativité des échantillons.

On contourne ces difficultés à l'aide de méthodes diverses qui permettent d'identifier un échantillon représentatif tout en fournissant des indications pour évaluer les effets de l'erreur échantillonnale, en fin de compte inévitable : on sait qu'il n'existe en effet aucune méthode d'échantillonnage capable de garantir la représentativité parfaite d'un échantillon.

Pour améliorer la représentativité d'un échantillon, il faut tout d'abord s'attacher à définir correctement les objectifs de la recherche et la population ; ce faisant, on assure une meilleure validité externe de la recherche.

1. Définition des objectifs de la recherche

L'importance de la définition des objectifs d'une recherche n'est plus à démontrer, d'autant plus que ces derniers conditionnent le choix de la population, de l'unité échantillonnale, du type et de la qualité des informations recueillies. Un ensemble d'objectifs énoncés dans une optique opérationnelle précisera la nature et la source des données requises et garantira que la totalité des informations nécessaires soient recueillies et que toute redondance soit éliminée. À titre d'exemple, imaginons que le ministère de l'Éducation du Québec décide de consulter l'opinion publique au sujet de la restructuration du système scolaire : les objectifs de cette recherche devront identifier, entre autres, les organismes ou groupes de citoyens visés, car il est bien évident que les parents, les enseignants, les élèves et les divers groupements religieux ou linguistiques fourniront des informations tout à fait différentes.

2. La définition de la population

Il est impossible de définir un échantillon valable sans connaître les caractéristiques de la population cible. Une population est un ensemble dont les éléments

comportent une ou plusieurs caractéristiques communes. À vrai dire, c'est l'ensemble des mesures qualitatives ou quantitatives de ces caractéristiques qui définit la population; les individus ou entités appartenant à la population ne sont que les porteurs des caractéristiques en question. Un sujet peut donc faire partie de plusieurs populations distinctes selon les caractéristiques retenues. Par exemple, supposons que l'individu dénommé Jean-Claude appartient à la population des « Montréalais », des « étudiants à l'Université de Sherbrooke », des « Canadiens d'origine haïtienne », des « mâles âgés de 20 à 25 ans » et des « joueurs de football » : dans une recherche sur le multiculturalisme, Jean-Claude pourrait être un élément acceptable de la population visée, mais sa caractéristique de « joueur de football » ne serait pas nécessairement pertinente.

Pour définir une population, il faut d'abord déterminer une règle qui permet d'identifier les éléments ou unités échantillonnales de la population; en outre, le choix d'un échantillon optimal implique qu'on en spécifie la structure interne, puisque ce sont les propriétés de cette structure qui conditionnent le choix d'une méthode d'échantillonnage susceptible d'augmenter la représentativité de l'échantillon retenu.

La formulation d'une règle d'identification des éléments d'une population et la détermination de sa structure ne vont pas sans difficulté. Pour revenir à notre exemple, supposons que le ministère de l'Éducation du Québec envisage d'étudier l'attitude des Montréalais à l'égard d'un projet de réforme scolaire. À première vue, la définition de la population cible pourrait sembler tout-à-fait claire et évidente: mais examinons-la un peu plus attentivement.

- a) Qui sont les « Montréalais » en question? S'agit-il des personnes
 - dont le lieu de résidence est Montréal?
 - qui travaillent à Montréal?
 - qui ont vécu une partie importante — mais laquelle? — de leur vie à Montréal?
 - qui correspondent à l'ensemble de ces caractéristiques, ou seulement à une combinaison particulière d'entre elles?
 - etc.?
- b) Quelle est la délimitation territoriale de l'entité « Montréal »?
 - S'agit-il
 - de la municipalité?
 - de l'Île?
 - du Montréal métropolitain?
 - etc.?
- c) Comment définir un « lieu de résidence »? S'agit-il
 - d'une maison dont on est propriétaire? d'un appartement loué? d'un emplacement dans un établissement public quelconque (hôpital, caserne, etc.)?

d) Qui peut être considéré comme « résident » ?

- le citoyen ? tout Canadien ou immigrant reçu ? toutes les personnes vivant dans une même résidence ?

Par ailleurs, faut-il retenir d'autres paramètres, se rapportant par exemple aux groupes d'âge, à la religion, etc. ?

On voit qu'à défaut d'apporter des réponses précises à ces questions, il est impossible d'établir une règle d'identification des éléments de la population visée. Quant à la structure interne de cette dernière, dont dépend, on s'en souvient, la détermination d'une méthode d'échantillonnage appropriée, il s'agit d'une question relativement complexe dont l'étude doit faire appel aux expériences acquises en la matière (les études antérieures, les cartes démographiques, etc.), sinon à une pré-expérimentation en bonne et due forme. Il faut au moins répondre le plus objectivement possible à des questions telles que :

- a) La population est-elle homogène par rapport à la (aux) caractéristique(s) étudiée(s) ?
- b) Dans le cas d'une population hétérogène, existe-t-il des groupes homogènes au sein de la population ?
- c) Un regroupement géographique ou administratif est-il souhaitable ?
- d) Quelle est l'attitude générale de la population face à la recherche ? Quel est son degré de collaboration ? etc. ?

Bref, on aura compris que la définition de la population cible mérite une attention toute particulière. Clarifions un dernier point : on désigne la population d'une manière assez vague sous l'appellation d'« univers » : Kerlinger (1973, p. 118), De Landsheere (1975, p. 251) et Dixon et Massey (1969, p. 37) en fournissent des exemples. Par contre, Cronbach *et al.* (1972, p. 23) parlent d'« univers » pour désigner le contenu d'un enseignement ou d'un objet psychologique ; chez Warwick et Lininger (1975, p. 71), ce même terme désigne « les événements ou les objets qui n'ont pas de limite numérique ». Compte tenu de ces variations de sens, l'emploi de la notion d'« univers » en tant que synonyme de « population » est à déconseiller.

II- Les méthodes d'échantillonnage

Nous avons déjà indiqué qu'en sciences humaines la complexité de la population est parfois tellement grande que le choix d'un échantillon adéquat s'avère une tâche quasi impossible. Ces difficultés sont à l'origine de la théorie et des techniques d'échantillonnage. Selon De Landsheere, « échantillonner, c'est choisir un nombre limité d'individus, d'objets ou d'événements dont l'observation permet de tirer des conclusions (*inférences*) applicables à la population entière à l'intérieur de laquelle le choix a été fait » (De Landsheere, 1975, p. 251). Les techniques d'échantillonnage constituent des moyens d'évaluer la qualité de ces inférences. Rappelons, encore une fois, qu'aucune technique d'échantillonnage ne peut garantir la représentativité

de l'échantillon choisi. Cependant, son degré de représentativité peut être évalué à l'aide de deux approches différentes. Selon que l'on tente d'établir cette représentativité de façon indirecte ou directe, on utilisera :

- a) des méthodes d'échantillonnage probabilistes, selon lesquelles le degré de représentativité augmente en même temps que diminue l'erreur échantillonnale des statistiques estimant les paramètres (de la population) ; ainsi, la validité des inférences est vérifiable au sens probabiliste du terme ;
- b) des méthodes d'échantillonnage non probabilistes englobant diverses techniques qui postulent ou établissent directement la représentativité d'un échantillon ; la validité des inférences n'est pas objectivement vérifiable, si ce n'est par l'accumulation des preuves au fil des années.

1. L'échantillonnage probabiliste

Toutes les techniques d'échantillonnage probabiliste sont fondées sur la méthode d'échantillonnage dite « aléatoire simple » (A.S.). Cette méthode accorde une « chance » (probabilité) de sélection égale et non nulle à tous les échantillons d'une grandeur donnée. Soit, pour fins d'illustration, une population constituée de cinq éléments dénotés A, B, C, D, et D, dont deux appartiennent à la même catégorie. Supposons qu'on veuille en extraire un échantillon de grandeur 2 sans remise³. Les dix échantillons possibles sont les combinaisons de cinq éléments pris deux à la fois, soit : (A, B), (A, C), (A, D), (A, D), (B, C), (B, D), (B, D), (C, D), (C, D) et (D, D) ; chacun de ces dix échantillons a une probabilité de sélection fixe de $1/10 = 2(1/5 \cdot 1/4)$. Il est évident que ces échantillons n'ont pas tous le même degré de représentativité : ainsi, (D, D) ne tient pas compte des éléments A, B et C, etc.

La méthode d'échantillonnage dite « aléatoire stratifiée » (A. St.) est un regroupement de plusieurs échantillonnages aléatoires simples. En fait, si la population est formée de plusieurs sous-populations (strates) distinctes telles que la variabilité au sein d'une même strate est minime par rapport à la variabilité entre les différentes strates, alors on extrait de chaque strate un sous-échantillon aléatoire simple dont le regroupement constitue l'échantillon final. L'efficacité⁴ de l'échantillonnage aléatoire stratifié est fonction du choix des variables de stratification : plus ces variables sont reliées aux variables étudiées, plus la stratification est efficace⁴. On voit clairement ici combien l'étude de la structure interne de la population est essentielle à l'efficacité de l'échantillonnage.

Sous réserve que les éléments de la population soient énumérés dans un ordre aléatoire, la méthode d'échantillonnage « systématique » est équivalente à l'échantillonnage aléatoire simple, et a l'avantage d'être facile à appliquer.

L'échantillonnage par « grappes » est une technique dont l'unité échantillonnale est un ensemble comprenant plus d'un élément de la population au lieu d'un seul, comme dans les autres méthodes. Cette technique est en fait un échantil-

lonnage aléatoire simple où l'unité d'échantillonnage est définie par un groupe d'éléments. L'emploi de ce type d'échantillon est très répandu en éducation, où les groupes de sujets sont souvent établis préalablement à la recherche ; on pense ici aux groupes « intacts » (par exemple, les classes d'une école).

L'échantillonnage en grappes pose habituellement un problème lié au fait que l'unité d'échantillonnage initiale est implicitement redéfinie lors de l'analyse et de l'interprétation des données ; or, cette redéfinition de l'unité n'est pas recevable. Supposons qu'on choisisse aléatoirement 15 classes parmi les 300 que comprend une population donnée ; la taille de l'échantillon est donc égale à 15, mais au moment de l'analyse la grandeur de l'échantillon est redéfinie incorrectement comme étant le nombre total d'élèves de ces 15 classes. Une telle confusion a pour effet d'augmenter la probabilité de résultats « significatifs » (!), mais il est clair que seule l'inférence reliée aux classes (plutôt qu'aux élèves) est légitime.

Il existe plusieurs autres techniques d'échantillonnage probabiliste. Parmi les développements plus récents, soulignons quelques apports notables dans les domaines de l'inférence par le plan d'échantillonnage (I-Plan) et par un modèle statistique (I-Mod), l'utilisation des variables auxiliaires, les problèmes créés par la non-réponse, etc. Le lecteur intéressé pourra consulter divers ouvrages consacrés à la théorie et à l'application des méthodes d'échantillonnage (voir, par exemple, Cochran, 1977 ; Raj, 1968, 1972 ; Warwick et Lininger, 1975 ; Krewski et al., 1981 ; Särndal, 1982).

2. L'échantillonnage non probabiliste

Cette catégorie comprend toutes les méthodes d'échantillonnage où aucune probabilité de sélection n'est assignée aux éléments de la population. Il en découle que les inférences établies à partir des résultats obtenus relèvent plutôt de l'intuition que d'une démarche objective et scientifique. Parmi ces techniques, citons les échantillonnages accidentel, intentionnel, par expert, par quota, etc. Selon ces méthodes, la détermination d'un échantillon dit « représentatif » dépend fondamentalement de l'accessibilité des éléments échantillonnés, de l'habileté ou de l'expertise du chercheur.

Un défaut majeur et fréquent des recherches qui ont recours à ces techniques consiste à interpréter les données à l'aide de modèles statistiques. Alors que tous les modèles de statistiques inférentielles sont fondés sur la notion d'échantillonnage probabiliste, leur utilisation dans le cadre d'échantillons non probabilistes est tout à fait absurde. Si un chercheur croit pouvoir affirmer la représentativité de ses échantillons, il est logique qu'il interprète les résultats obtenus (les statistiques provenant des échantillons) comme des paramètres de la population correspondante. Ses résultats seraient alors logiquement significatifs.

Un autre problème résulte d'un compromis élégant par lequel on essaie de justifier l'utilisation des échantillons non probabilistes en généralisant les résultats selon les règles des statistiques inférentielles. C'est la solution dite de la « population inconnue » selon laquelle tout échantillon est potentiellement l'échantillon aléatoire d'une population inconnue; par conséquent, les résultats obtenus sont théoriquement généralisables à cette même population. Mais il faut s'interroger sur l'importance pratique de ce type d'inférence. Peut-on régler un problème pour une population inconnue, c'est-à-dire non définie? Dans un pareil cas il vaut mieux tout simplement admettre que les résultats sont valables uniquement pour l'échantillon choisi.

Exposons enfin un dernier point de vue, malheureusement trop répandu et qui relève d'une autre forme de « compromis élégant », par lequel le chercheur insiste sur le fait qu'en pratique le choix d'un échantillon « strictement » aléatoire est impossible et que tous les échantillons ne seraient en réalité que des échantillons non probabilistes; il n'y aurait d'autre partie d'alternative que de se plier à la réalité en s'accommodant d'échantillons non probabilistes. Certes, il faut admettre que le choix d'un échantillon aléatoire pose un problème réel; par ailleurs la théorie de l'échantillonnage est un modèle dont il n'est pas toujours possible de satisfaire toutes les exigences. Mais il ne faut pas en conclure pour autant que ce modèle peut être ignoré. Au contraire, la présence éventuelle d'erreurs dues à l'écart entre le modèle et la réalité devrait inciter le chercheur à réduire cet écart dans toute la mesure du possible, de façon à garantir une meilleure validité externe des résultats obtenus.

III- L'estimation de la taille d'un échantillon de sujets

Un échantillon trop petit diminue la pertinence des résultats d'une recherche; par contre, un échantillon trop grand entraîne un gaspillage de ressources. Quelle est la grandeur de l'échantillon qui permette d'atteindre une précision donnée ou, pour un coût donné, d'atteindre le maximum de précision? Bien qu'il n'y ait pas de réponses simples à cette question, on pourra s'inspirer de la démarche suivante pour déterminer la taille d'un échantillon donné (voir, à ce sujet, Cochran, 1977, chapitre 4):

- a) Expliciter les fins pour lesquelles les données sont recueillies de même que toute décision susceptible d'en découler. À cet égard, il est nécessaire de déterminer le degré minimal de précision que doivent atteindre les résultats pour être utilisables.
- b) Établir une équation qui relie la grandeur de l'échantillon à la précision exigée et aux restrictions imposées. De telles équations sont habituellement disponibles pour les méthodes d'échantillonnage probabilistes, et c'est là un avantage certain.

- c) Estimer les paramètres de l'équation mentionnée en b) à partir des résultats de recherches analogues ou d'une préexpérimentation.

Soulignons enfin que le problème de l'échantillonnage des sujets ne couvre qu'une partie du problème général de l'échantillonnage. En effet, les instruments de mesure utilisés dans le cadre d'une recherche particulière ne représentent qu'un échantillon de tous les instruments potentiels, tout comme les conditions expérimentales ne sont que des échantillons de toutes les conditions expérimentales possibles. L'erreur de mesure ainsi que d'autres erreurs dites non échantillonnales jouent un rôle fort important dans les développements méthodologiques récents; l'intégration des méthodes de contrôle des erreurs tant échantillonnales que non-échantillonnales pourrait jeter les bases d'une nouvelle discipline dénommée *survey research* (Dalenius, 1981; Namboodiri, 1978). Au cours du présent article nous n'avons abordé que le problème de l'échantillonnage des sujets; le lecteur pourra consulter l'excellent volume de Cronbach *et al.* (1972) sur la théorie de la généralisabilité qui traite des différentes sources d'erreur de mesure.

Nous nous sommes limités jusqu'ici aux éléments les plus importants du problème de l'échantillonnage qui sont directement reliés à la validité externe de la recherche en sciences de l'éducation et, de façon plus générale, en sciences humaines. Dans ce qui suit, on présentera les résultats d'une recherche (Gougeon, 1982) consacrée au problème de l'échantillonnage des sujets dans les thèses de doctorat et les mémoires de maîtrise des étudiants en psychologie et en sciences de l'éducation de deux universités québécoises. L'objectif de cette étude n'est pas de comparer les universités ou les entités administratives concernées, mais plutôt, à partir d'un exemple illustrant des lacunes et des difficultés réelles, de sensibiliser les éducateurs responsables de la formation des chercheurs dans ces deux disciplines aux problèmes de l'échantillonnage.

IV- La problématique de la recherche

Une étude préliminaire de quelques mémoires et thèses universitaires en sciences de l'éducation et en psychologie, jointe à une analyse des écrits, a permis d'identifier les problèmes plus spécifiques de la recherche en sciences humaines:

- a) le problème de la définition de la population;
- b) le choix d'un échantillon représentatif;
- c) le phénomène de mortalité;
- d) les contraintes d'ordre pratique;
- e) le choix d'un échantillon occasionnel et la définition d'une population de référence;
- f) les limites des inférences.

Les résultats de cette étude préliminaire ont servi de toile de fond à la recherche qui se proposait de répondre à deux questions:

- «Quelle est l'étendue réelle du problème de l'échantillonnage tel qu'il se manifeste dans les mémoires et thèses présentés par des chercheurs en sciences de l'éducation et en psychologie de deux universités québécoises?»
- «Quelles solutions peut-on proposer pour y remédier?»

Par ailleurs, cette étude préliminaire a permis de déterminer les facteurs pertinents suivants:

- a) la définition de la population;
- b) le choix du plan d'échantillonnage:
 - degré de précision désiré et choix de la taille de l'échantillon;
 - description de la technique d'échantillonnage;
 - rationnel pour le choix du plan échantillonnal;
- c) les inférences:
 - plausibilité;
 - limites, biais, erreurs.

Notons tout de suite que des contraintes de toutes sortes ont limité l'ampleur de l'étude. Les deux universités retenues ont été choisies en raison de leur accessibilité; il n'est donc pas question de généraliser les résultats.

V- Le schéma expérimental

Les variables d'intérêt de la recherche sont: les techniques d'échantillonnage et d'inférence utilisées dans les mémoires et les thèses acceptés en 1978 et en 1979 en éducation et en psychologie par deux universités québécoises, désignées ci-après par les lettres «A» et «B».

La technique d'échantillonnage retenue est de nature probabiliste; il s'agit de l'échantillonnage aléatoire stratifié proportionnel du cinquième de la population. La population a été subdivisée en huit strates selon trois variables de stratification, chacune possédant deux caractères de classification: les universités (A et B), les disciplines (éducation et psychologie) et les niveaux (maîtrise et doctorat). La stratification a été effectuée par souci de représentativité et non dans le but de comparer les universités ou les départements; cependant, les niveaux de maîtrise et de doctorat seront comparés. La fraction échantillonnale (un cinquième) a été déterminée en imposant une erreur-type de 0,05 dans l'estimation de nos paramètres. Ce niveau de précision a semblé suffisant pour répondre aux objectifs de la recherche. Le tableau 1 indique la taille des différentes strates de la population, alors que le tableau 2 indique la taille des échantillons correspondants.

À la suite d'une première expérimentation, une grille d'observation⁶ répondant aux objectifs de la recherche a été établie afin d'évaluer tout d'abord la qualité des rapports étudiés du point de vue des variables retenues (concernant les populations, les échantillons et les inférences) et, ensuite, de compiler les défauts observés. La grille permet de répondre aux questions suivantes:

Tableau 1
Taille de la population

	Maîtrise		Doctorat		Total	
Éducation	A	69	A	26	A	95
	B	6(*)	B	0	B	6
	E	75	E	26	E	101
Psychologie	A	253	A	37	A	290
	B	94	B	0	B	94
	E	347	E	37	E	384
Total	A	322	A	63	A	385
	B	100	B	0	B	100
	E	422	E	63	E	485

(*) A: l'université A;
 B: l'université B;
 E: l'ensemble des universités A et B.

Tableau 2
Taille des échantillons

	Maîtrise		Doctorat		Total	
Éducation	A	14	A	6	A	20
	B	6	B	0	B	6
	E	20	E	6	E	26
Psychologie	A	51	A	8	A	59
	B	19	B	0	B	19
	E	70	E	8	E	79
Total	A	65	A	14	A	79
	B	25	B	0	B	25
	E	90	E	14	E	104

- a) La population est-elle définie correctement?
- b) Le choix de la population est-il justifié?
- c) Le choix du plan échantillonnal est-il justifié?
- d) Le choix de la taille de l'échantillon est-il justifié?
- e) Quel est le type d'échantillonnage utilisé?
- f) La méthode d'échantillonnage est-elle décrite adéquatement?
- g) Le plan échantillonnal permet-il de généraliser les résultats obtenus?
- h) Les résultats sont-ils généralisés?
- i) Est-il fait mention des limites, des biais ou des erreurs?

VI- *Le schéma de l'analyse*

Dans le but d'obtenir des informations relatives à:

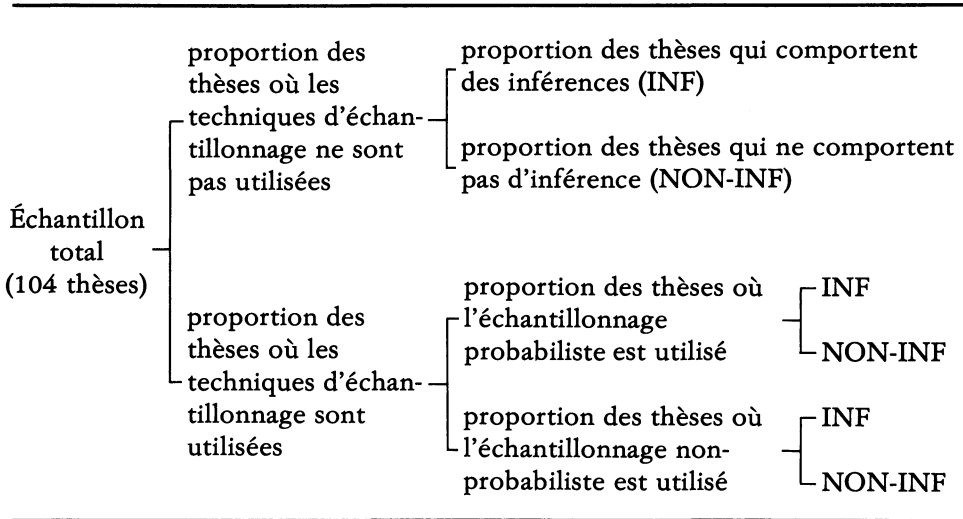
- a) l'étendue de l'utilisation des techniques d'échantillonnage dans la population étudiée;
- b) l'importance des techniques d'échantillonnage probabiliste et non probabiliste;
- c) l'étendue des défauts observés;
- d) l'amélioration de la qualité de l'échantillonnage des thèses de troisième cycle par rapport aux mémoires de deuxième cycle,

l'analyse des données a porté sur:

- a) l'estimation de la proportion des thèses et mémoires où une technique d'échantillonnage est utilisée;
- b) l'estimation de la proportion des thèses et mémoires utilisant des techniques probabilistes et non probabilistes, ainsi que de la proportion des thèses et mémoires qui comportent des inférences exactes;
- c) l'estimation de la proportion des thèses et mémoires présentant des défauts dénotés D_1, D_2, \dots, D_8 , concernant:
 - la définition de la population (D_1),
 - la justification de la population (D_2),
 - la justification du plan échantillonnal (D_3),
 - la justification de la taille de l'échantillon (D_4),
 - la spécification et la description de la méthode d'échantillonnage (D_5),
 - la possibilité de réaliser une inférence exacte (D_6),
 - l'exactitude de l'inférence réalisée (D_7),
 - les biais et les limites du plan échantillonnal (D_8).

Le diagramme 1 résume le schéma de l'analyse des résultats observés pour l'ensemble de l'échantillon.

Diagramme 1
Schéma de l'analyse pour l'échantillon total



VII- L'analyse et l'interprétation des résultats

Dans ce qui suit, les résultats ont été regroupés en cinq ensembles. Toutes les estimations ont été réalisées à l'aide d'intervalles de confiance de 95% ($\alpha = 0,05$).

1. Il a été estimé qu'entre 62,40% et 77,98% des thèses⁷ de la population visée utilisent une technique d'échantillonnage. Ce résultat indique que la pratique de l'échantillonnage est suffisamment répandue et justifie l'intérêt porté aux problèmes qu'elle pose. Pour la suite des analyses, la population cible se réduit à la sous-population des thèses où les techniques d'échantillonnage sont utilisées. En ce qui concerne les autres thèses, seule l'étude des inférences inappropriées a été réalisée.
2. Il a été estimé qu'entre 64,91% et 83,03% des thèses de la sous-population mentionnée ci-haut utilisent des méthodes d'échantillonnage non probabiliste. Le tableau 3 complète l'information concernant le type d'échantillonnage utilisé. On pourrait s'interroger sur les raisons qui amènent les chercheurs à ne pas choisir en plus grand nombre une technique probabiliste. Selon nous, une bonne part de l'explication de ce phénomène réside dans les contraintes physiques propres à la recherche en sciences humaines. À ce chapitre, notons, entre autres, le refus de coopérer, la nécessité quasi absolue, lorsque la recherche porte sur des groupes spéciaux, de choisir un groupe intact constituant un échantillon disponible ou jugé représentatif par un expert, etc. Il faut aussi considérer que les

Tableau 3
Répartition des thèses selon le type d'échantillonnage utilisé
 ($\alpha = 0,05$)

Type	Taille	Proportion	Intervalle de confiance
Probabiliste	19	0,2603	(0,1697 , 0,3508)
Non probabiliste	54	0,7397	(0,6491 , 0,8303)

étudiants de maîtrise et de doctorat (les auteurs des recherches étudiées) ne disposent pas toujours des ressources nécessaires à la détermination d'un échantillon selon les techniques probabilistes. Enfin, on ne pourrait passer sous silence la possibilité que la formation de ces étudiants présente des lacunes au niveau des méthodes d'échantillonnage. Ces raisons nous amènent à formuler l'hypothèse, non confirmée dans la présente étude, que l'utilisation des méthodes d'échantillonnage non probabiliste ne relève pas toujours d'un choix bien réfléchi mais pourrait être le deuxième choix d'un chercheur peu ou mal renseigné sur les techniques d'échantillonnage probabiliste et qui chercherait un refuge illusoire dans des approches moins « techniques ».

3. En ce qui concerne la même sous-population (des thèses où les techniques d'échantillonnage sont utilisées), il a été estimé qu'entre 86,11% et 97,45% des thèses de cette population comportent au moins un des défauts énumérés à la section 6. Ce résultat est tout à fait inquiétant et justifie quelques interrogations. Le tableau 4 présente la liste des défauts (D_1 , ..., D_8) ainsi que leur fréquence d'occurrence. Aucune des proportions des défauts observés ne peut être considérée comme négligeable.

Les défauts le plus souvent observés sont les suivants: l'absence de généralisabilité des résultats (D_6), l'absence de justification de la taille échantillonnale (D_4), l'insuffisance de la description de la méthode échantillonnale (D_5). Étant donné la forte proportion des thèses où l'échantillonnage non probabiliste est utilisé (entre 64,91% et 83,03%), la fréquence élevée des défauts D_4 et D_6 était à prévoir. Par contre, il est difficile d'expliquer la fréquence élevée du défaut D_5 : quelle que soit la méthode d'échantillonnage retenue, rien n'empêche de la décrire adéquatement. Ce défaut est d'autant plus injustifié qu'il rend impossible toute répétition des études, processus essentiel à la recherche scientifique.

Les défauts moins fréquents n'en demeurent pas moins graves. Ainsi, entre 23,18% et 42,58% des thèses ne font aucune mention de limites, d'erreurs ou de biais du plan échantillonnal (D_8). Pourtant ces données sont essentielles pour

Tableau 4
Proportion des thèses ayant le défaut D_i , où i varie de 1 à 8
($\alpha = 0,05$)

Défauts concernant	Taille	Proportion	Intervalle de confiance
D ₁ : la définition de la population	18	0,2466	(0,1576 , 0,3356)
D ₂ : la justification de la population	17	0,2329	(0,1456 , 0,3202)
D ₃ : la justification du plan échantillonnal	21	0,2877	(0,1942 , 0,3812)
D ₄ : la justification de la taille de l'échantillon	39	0,5342	(0,4312 , 0,6372)
D ₅ : la description de la méthode d'échantillonnage	28	0,3836	(0,2832 , 0,4840)
D ₆ : la possibilité de réaliser une inférence	57	0,7808	(0,6934 , 0,8682)
D ₇ : l'exactitude de l'inférence réalisée	23	0,3151	(0,2192 , 0,4110)
D ₈ : la mention des limites	24	0,3288	(0,2318 , 0,4258)

juger de la valeur des conclusions tirées. Entre 21,92% et 41,10% des cas observés comportent des inférences qui dépassent les limites de leur plan échantillonnal (D₇). Là encore, ce défaut touche à un aspect fondamental: la connaissance qu'a le chercheur des limites de sa recherche.

Enfin, entre 15,76% et 33,56% des thèses comportent des définitions de population incorrectes (D₁), et entre 14,56% et 32,02% ne présentent aucune justification du choix de la population (D₂). À ce sujet, rappelons que la définition de la population est l'une des premières étapes à réaliser lors de l'élaboration d'une recherche. Autrement, comment pourrait-on choisir un plan échantillonnal adéquat, c'est-à-dire qui optimise la représentativité de l'échantillon si la population qu'il représente est inconnue?

Parmi les résultats les plus importants, notons que 10,96% des cas observés, à quoi correspond un intervalle de confiance de (4,51%, 17,41%), réunissent *tous* les défauts. Notons aussi que 47,95% des cas observés comportent au moins trois défauts. S'il peut sembler raisonnable de tolérer un ou deux défauts jugés secondaires, bien qu'une thèse devrait en être exempte, nous ne pouvons éviter de nous interroger sur la crédibilité à accorder aux résultats de telles études et, vu

la nature prétendûment formative de tels travaux, nous demander comment de telles lacunes ont pu passer inaperçues au moment de leur évaluation!

4. En ce qui concerne la sous-population des thèses où une technique d'échantillonnage probabiliste a été utilisée, il a été établi :
 - a) que les thèses à échantillonnage probabiliste présentent beaucoup moins de défauts que les thèses à échantillonnage non probabiliste;
 - b) que la proportion des thèses à échantillonnage probabiliste ayant au moins trois défauts est inférieure à celle des thèses à échantillonnage non probabiliste;
 - c) que tous les cas entièrement corrects (ne présentant aucun défaut) se trouvent dans la catégorie à échantillonnage probabiliste.

Ces résultats peuvent s'expliquer du fait que, par leur nature, certains défauts affectent principalement les thèses où une technique d'échantillonnage non probabiliste est utilisée. Le choix d'une technique non probabiliste se trouve ainsi lié à des défauts indicateurs d'un manque de planification dans la conception du schéma expérimental.

5. Finalement, l'étude a établi que les thèses de doctorat ne présentent aucune amélioration par rapport aux mémoires de maîtrise. Cet état de chose vient renforcer la constatation, émise à la section 7.2, concernant les lacunes au niveau de la formation des étudiants.

VIII- Les limites de la recherche

On abordera ici l'étude des facteurs pouvant affecter la validité des conclusions: l'erreur échantillonnale, les biais de sélection, les biais d'estimation, et l'erreur de mesure.

1. Erreur échantillonnale

L'erreur-type des estimateurs des proportions a été fixée à 0,05. Ceci a conduit à choisir une grandeur échantillonnale⁸ d'environ 100 thèses; il en résulte des intervalles de confiance (à 95%) du type ($P - 0,10$, $P + 0,10$), où P est une proportion. Il n'a pas paru nécessaire de rechercher une plus grande précision étant donné les objectifs de notre recherche, qui étaient de vérifier l'existence du problème autant que sa nature.

2. Biais de sélection

Il n'existe pas à proprement parler de biais de sélection dans le type des recherches réalisées. Tous les éléments de la population ont une probabilité égale d'être choisis. De même, il ne peut y avoir de phénomène de mortalité échantillonnale puisque tous les documents sont disponibles.

3. *Biais d'estimation*

Lors du calcul de l'intervalle de confiance pour les sous-populations étudiées, la taille des sous-populations a été estimée en prenant la valeur maximale des intervalles de confiance pour la population totale. Les intervalles de confiance sont alors plus grands qu'en réalité, mais les estimations choisies, P_i , ne sont pas biaisées par rapport aux paramètres.

4. *Erreur de mesure*

L'erreur de mesure est sans conteste le genre d'erreur le plus souvent présent dans ce type de recherche. C'est pourquoi la mise au point de l'instrument de mesure a été réalisée avec soin : l'instrument a été révisé à deux reprises avant son utilisation. La grille⁶ d'observation qui en résulte devrait permettre de classer le plus objectivement possible les observations.

Conclusion

Nous avons tenté dans cet article de clarifier le lien entre la théorie de l'échantillonnage et la validité externe d'une certaine catégorie de recherches en sciences humaines. La propriété de généralisabilité des résultats concernant les sujets, c'est-à-dire l'extension des conclusions d'une recherche à un ensemble plus vaste qui englobe le groupe étudié, dépend directement d'éléments fondamentaux de la théorie de l'échantillonnage. Ces éléments ont trait, d'une part à la définition des objectifs de la recherche et à la définition de la population étudiée et, d'autre part, aux méthodes d'échantillonnage. Ces premiers aspects constituent l'infrastructure de toute inférence qui conditionnera les aspects suivants, lesquels permettront d'évaluer la qualité de ces inférences. Selon que l'évaluation se fait directement ou indirectement, on utilise des méthodes d'échantillonnage non probabilistes ou probabilistes. Il existe une confusion généralisée sur la nature des inférences possibles à l'aide de ces deux familles de méthodes. Alors que les méthodes non probabilistes fondent les inférences sur des déductions « logiques », les méthodes probabilistes établissent les inférences à partir d'une interprétation « statistique » des résultats. L'incompatibilité des deux approches est à l'origine d'une source majeure d'inférences incorrectes. Tout se passe comme si certains chercheurs tentaient, par un « compromis élégant », d'obtenir le meilleur des deux mondes : éviter les exigences techniques de l'approche probabiliste tout en bénéficiant de l'aura de scientificité des inférences statistiques. Quant à l'éternelle question de la taille de l'échantillon, question apparemment banale et « technique », il s'agit de la pointe de l'iceberg qui cache toute la question de la validité externe de la recherche en sciences humaines, la justification de la taille échantillonnale étant étroitement reliée aux objectifs de la recherche, à la définition de la population, au plan échantillonnal, au type et à la qualité de l'inférence désirée.

Il restait à démontrer l'existence du problème, sinon dans une population d'envergure regroupant toutes les catégories de la recherche en sciences humaines — tâche colossale, voire impossible —, du moins dans un ensemble plus réduit mais quand même important. Le choix des thèses en psychologie et en éducation a été essentiellement motivé par le caractère formateur de tels travaux et par les contrôles existants, supposément garants de qualité. L'existence des problèmes de validité externe a été vérifiée et leur nature, connue. Que l'envergure des défauts atteigne des proportions épidémiques souligne le sérieux de la situation et soulève à son tour le problème de la source de ces difficultés, de ses conséquences et des solutions possibles. Ce problème nécessite donc des études plus approfondies. Rappelons que l'objectif principal de cette étude est de sensibiliser les milieux éducatifs aux problèmes de l'échantillonnage. Les auteurs espèrent qu'elle suscitera suffisamment d'intérêt pour inciter à poursuivre la recherche et à en arriver aux correctifs nécessaires. En attendant, il a paru utile d'énoncer les quelques remarques suivantes qui pourraient servir de prémisses à des études ultérieures.

Malgré que certaines faiblesses observées dans les thèses étudiées, telle la profusion des techniques d'échantillonnage non probabiliste incorrectement appliquées, peuvent dépendre des contraintes matérielles inhérentes aux ressources réduites des étudiants, il n'en demeure pas moins que les résultats obtenus semblent indiquer un manque de préparation généralisée à la recherche. Notons à ce sujet qu'il n'existe pas à proprement parler de cours sur les méthodes d'échantillonnage dans les entités administratives responsables de la formation des étudiants dont nous avons étudié les thèses. Si le sujet est abordé dans certains cours de méthodologie de la recherche, à en juger par les résultats il doit s'agir d'un traitement superficiel certainement insuffisant. Par ailleurs, certaines contraintes administratives rencontrées dans les écoles du système scolaire québécois ont un impact négatif sur la qualité de la validité externe des recherches qui y sont faites. À ce propos, on ne peut que déplorer l'absence d'un réseau complet d'écoles pilotes où, en étroite collaboration avec les universités, l'étudiant pourrait mener sa recherche dans des conditions contrôlées. Finalement, les directeurs de recherche et les jurys concernés doivent orienter leurs étudiants en ce qui concerne les méthodes d'échantillonnage d'une manière plus adéquate et assumer leurs responsabilités dans ce domaine. Le fait qu'une infime proportion des thèses soit jugée satisfaisante présente un double danger : celui pour le lecteur averti de considérer l'ensemble des recherches en sciences humaines comme non valable, et pour le lecteur non averti d'accepter les conclusions incorrectes sans discernement.

NOTES

1. À ce sujet, on pourra consulter Cochran (1977).
2. Les concepts de validité interne et de validité externe sont exposés, entre autres, par Kerlinger (1973) à la suite de Campbell et Stanley (1963), et de Cook et Campbell (1979).

3. Pour un échantillon donné, un élément sélectionné sans remise ne peut être replacé dans la population, et, par conséquent, ne peut être sélectionné plus d'une fois.
4. Dans ce contexte, les termes « efficacité » et « efficace » ne sont pas utilisés dans leur acception technique exacte mais désignent plutôt le degré d'amélioration de la représentativité de l'échantillon obtenu en passant de la méthode aléatoire simple à la méthode aléatoire stratifiée.
5. La précision d'une estimation est une mesure de la variabilité des estimations obtenues par la répétition de la méthode d'échantillonnage utilisée.
6. Les lecteurs intéressés pourront consulter la grille ainsi que son guide d'utilisation auprès des auteurs. En ce qui concerne la méthode de construction de la grille, on consultera Gougeon (1982).
7. Le terme « thèse » utilisé sans autre attribut désignera, dans ce qui suit, les mémoires de maîtrise aussi bien que les thèses de doctorat.
8. La taille a été établie en négligeant la correction pour une population finie, soit environ le cinquième de la taille de la population; l'estimation qui résulte est donc conservatrice.

RÉFÉRENCES

- Campbell, Donald, T. et Stanley, Julian, C., *Experimental and quasi-experimental designs for research*, Chicago: Rand McNally, 1963.
- Cochran, W.G., *Sampling Techniques*, (3^e éd.), New York: Wiley, 1977.
- Cook, Thomas, D. et Campbell, Donald, T., *Quasi-Experimentation*, Chicago: Rand McNally, 1979.
- Cronbach, Lee, J. et al., *The Dependability of Behavioral Measurement: Theory of Generalizability for Scores and Profiles*, New York: Wiley, 1979.
- Dalenius, T.E., The Survey Statistician's Responsibility for Both Sampling and Measurement Errors, dans *Current Topics in Survey Sampling*, New York: Academic Press, 1981, p. 17-29.
- De Landsheere, G., *Introduction à la recherche en éducation*, Paris: A.C.B., 1975.
- Dixon, W.J. et Massey, F.J., *Introduction to Statistical Analysis*, New York: McGraw-Hill, 1975.
- Gougeon, Hélène, *Problème de l'échantillonnage dans la recherche en sciences humaines: Éducation et psychologie*, Mémoire de maîtrise, Université de Montréal, Faculté des Sciences de l'Éducation, Section Mesure et Évaluation, 1982.
- Grand Larousse Encyclopédique*, tome 4, Paris: Librairie Larousse, 1961.
- Kerlinger, F.N., *Foundations of Behavioral Research*, (2^e éd.), New York: Holt, Rinehart & Winston, 1973.
- Krathwohl, D.R., Future of Educational Research, *Educational Researcher*, vol. 11, n° 8, p. 13-15, 1982.
- Krewski, D., Platek, R. et Rao, J.N.K., *Current Topics in Survey Sampling*, New York: Academic Press, 1981.
- Lindquist, E.F., *Statistical analysis in educational research*, Cambridge: The Riverside Press, 1940.
- McNamara, J.F., Determining sample size in decision-oriented research studies, *Planning and changing*, vol. 9, n° 2, p. 125-127, 1978.
- Namboodiri, N.R., (Ed), *Survey Sampling and Measurement*, New York: Academic Press, 1978.
- Raj, Des, *The design of sample surveys*, New York: McGraw-Hill, 1972.
- Raj, Des, *Sampling theory*, New York: McGraw-Hill, 1968.
- Ross, K., Sample design for education survey research, *Evaluation in education: international progress*, vol. 2, n° 2, p. 105-195, 1978.
- Särndal, C.E., L'inférence statistique et l'analyse des données sous des plans d'échantillonnage complexes, notes de conférences, Séminaire de mathématiques supérieures, Montréal, juillet-août 1982.
- Sukhatme, P.V., *Sampling theory of surveys with applications*, Iowa State University Press, 1970.
- Warwick, D.P. et Lininger, C.A., *The sample survey: Theory and practice*, New York: McGraw-Hill, 1975.
- Winer, B.J., *Statistical principles in experimental design*, New York: McGraw-Hill, 1962.
- Yates, F., *Sampling methods for censuses and surveys*, Londres: Charles Griffin & Co., 1960.