Didactique
Apprentissage
et enseignement

https://doi.org/10.37571/2022.0204
2022, Vol. 3, N. 2, pp. 103-121

Recruter de meilleurs échantillons en éducation

Éric Dion[1], Eugene Borokhovski[2], Stéphane Cyr[1], Isabelle Plante[1], Patrick Charland[1]
[1] *Université du Québec à Montréal, Québec, Canada*
[2] *Université Concordia, Québec, Canada*

Pour citer cet article:
Dion, É., Borokhovski, E., Cyr, S., Plante, I. et Charland, P. (2022). Recruter de meilleurs échantillons en éducation. *Didactique*, *3*(2), pp. 103-121.
https://doi.org/10.37571/2022.0204

Résumé : Pour générer des conclusions valables, les études en éducation doivent reposer sur des échantillons de qualité. En plus d'être d'une taille suffisante, ces derniers doivent être représentatifs, c'est-à-dire être une version miniature de la population. Le présent article méthodologique décrit les défis liés au recrutement d'un échantillon de qualité étant donné l'impossibilité de recruter purement au hasard. Cette impossibilité découle des règles éthiques régissant les sciences sociales en général, incluant la recherche en éducation et en didactique. Nous avançons que malgré son utilité, la stratification (ex. : représenter les filles et les garçons dans des proportions réalistes) ne garantit pas la représentativité. Nous faisons également valoir que l'impératif de reproductibilité (c.-à-d. de démontrer que d'autres échantillons mènent à des résultats similaires) ne dispense pas les chercheurs de constamment s'efforcer de recruter les meilleurs échantillons possibles. À cette fin, l'article propose une stratégie en deux volets qui consiste à 1) définir un bassin de recrutement idéal en fonction des objectifs de recherche et 2) minimiser le taux de refus au sein de ce bassin. En général, les chercheurs ne devraient pas recourir à des échantillons de convenance dont l'unique mérite est d'être facilement accessibles.

Mots-clés: échantillon, recrutement, biais, représentativité, stratification, reproductibilité

https://doi.org/10.37571/2022.0204
2022, Vol. 3, N. 2, pp. 103-121

Introduction

En éducation, les théories et recommandations pratiques s'appuient sur des données habituellement recueillies auprès d'échantillons (ex. : d'une centaine d'élèves). Le recours aux échantillons est nécessaire puisque les chercheurs disposent rarement d'une information suffisante sur la population, c'est-à-dire l'ensemble des personnes ou institutions d'intérêt. Dans la mesure où les échantillons sont des versions miniatures de la population, il est possible d'en tirer des conclusions d'une portée générale (pour un historique de cette idée, voir Zhao, 2021). À titre d'exemple, une pratique d'enseignement est considérée efficace pour la population d'élèves lorsqu'elle permet à un échantillon de ceux-ci de faire de meilleurs apprentissages (p. ex. : What Works Clearinghouse, 2021).

Étant donné le rôle central de l'échantillon en recherche, il est impératif d'accorder un grand soin à son recrutement. Autrement, la valeur et la portée de l'étude peuvent être irrémédiablement compromises. Imaginons un chercheur s'intéressant aux besoins des enseignants en matière d'utilisation du tableau blanc interactif. Si par inadvertance le chercheur recrute un échantillon formé en majorité d'enseignants atypiques (ex. : avec une formation initiale en informatique), les conclusions de son étude risquent de ne pas refléter la réalité de l'enseignant typique.

Une étude reposant sur un échantillon problématique pourrait néanmoins orienter la prise de décision. En effet, plusieurs acteurs insistent sur l'importance du transfert des connaissances issues de la recherche sans encourager les utilisateurs à se renseigner sur la méthodologie. Mentionnons par exemple les lignes directrices du Réseau d'information pour la réussite éducative (2021) pour la soumission d'articles diffusés sur son site. Ce réseau recommande de « présenter succinctement la méthodologie », « seulement si cela est nécessaire ». Puisque l'utilisateur n'est pas toujours en mesure de juger des limites d'une étude (ex. : parce qu'il n'a pas accès à l'information pertinente), c'est au chercheur qu'il incombe de spécifier clairement en quoi son échantillon limite la portée des conclusions de l'étude. Cela dit, l'idéal est évidemment d'utiliser un échantillon approprié. Notre objectif est de définir les caractéristiques des échantillons de qualité et de formaliser une stratégie réaliste pour recruter de tels échantillons. Bien que la composition des échantillons influence les conclusions de tous les types d'étude (Palinkas, Horwitz, Green, Wisdom, Duan et Hoagwood, 2015; Robinson, 2014), la terminologie et les concepts utilisés dans ce qui suit réfèrent à la recherche quantitative.

https://doi.org/10.37571/2022.0204
2022, Vol. 3, N. 2, pp. 103-121

La taille

Tel que nous le démontrons ci-dessous, la qualité d'un échantillon dépend de deux caractéristiques : sa taille et sa représentativité. Dans les sciences sociales comme l'éducation, la réflexion s'est centrée sur la première caractéristique, la taille, c'est-à-dire le nombre d'élèves, d'enseignants ou d'autres intervenants qui participent à l'étude.

À certaines conditions, plus un échantillon est de grande taille, plus il permet d'inférer avec une faible marge d'erreur les caractéristiques de la population (pour l'origine de cette notion, voir Neyman, 1934). Un grand échantillon est également avantageux parce qu'il confère une puissance aux analyses statistiques, c'est-à-dire qu'il leur permet de révéler des différences ou des liens subtils. Étant donné l'importance accordée à la taille de l'échantillon et à la puissance statistique par les méthodologues (ex. : Luo, Li, Baek, Chen, Lam et Semma, 2021), il serait tentant de penser qu'un échantillon de grande taille (ex. : de centaines voire de milliers d'élèves) permet nécessairement d'aboutir à des conclusions fiables. Ce n'est pourtant pas le cas.

Pour illustrer ce point, considérons ce qui peut se produire lorsque les chercheurs recourent aux services de la plate-forme de travail en ligne *Mechanical Turk* fondée en 2005 par l'entreprise *Amazon* pour faire réaliser des tâches informatiques qui ne peuvent pas être automatisées (Buhrmester, Talaifar et Gosling, 2018). Les travailleurs s'inscrivant sur la plate-forme, c'est-à-dire les participants potentiels, se font proposer différentes tâches (« Human Intelligence Tasks ») par le site, notamment répondre à des questions d'enquête, en échange d'une compensation habituellement modeste. Même si *Mechanical Turk* représente une façon rapide et économique de recruter de grands échantillons, il est raisonnable de s'interroger sur la pertinence d'une telle stratégie. Par exemple, dans leur enquête destinée aux parents, Jensen-Doss, Patel, Casline, Ringle et Timpano (2021) ont observé qu'un travailleur *Mechanical Turk* sur cinq avait répondu au questionnaire en moins de 60% du temps prévu, probablement sans vraiment lire les questions (voir aussi Aruguete, Huynh, Browne, Jurs, Flint et McCutcheon, 2019). Des chercheurs utilisant le service soupçonnent même que certains de leurs questionnaires ont été frauduleusement complétés par des algorithmes plutôt que par humains (Kennedy et al., 2018). La taille n'est pas donc garante, en soi, de qualité et les chercheurs devraient résister à la tentation de concevoir leur stratégie de recrutement avec seul objectif d'obtenir un grand échantillon.

https://doi.org/10.37571/2022.0204
2022, Vol. 3, N. 2, pp. 103-121

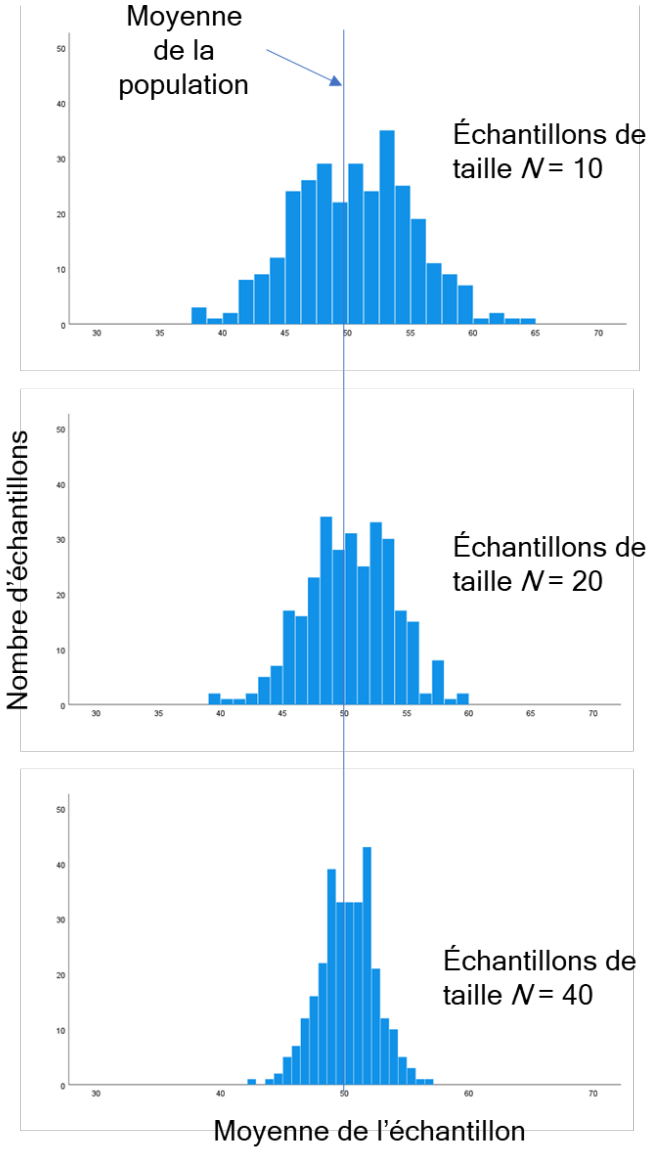
La représentativité et le recrutement au hasard

La représentativité est la deuxième caractéristique d'un échantillon de qualité. Un échantillon est représentatif lorsque sa composition ressemble à celle de la population (Berkeley Statistics, 2021). Recruter purement au hasard représente la meilleure stratégie pour maximiser la probabilité d'obtenir un tel échantillon. Lorsque cette stratégie est utilisée, c'est *exclusivement* le hasard qui détermine qui participera parmi les personnes répondant aux critères établis en fonction des objectifs de recherche. Aucune de ces personnes n'est exclues d'emblée (ex. : parce qu'elles travaillent en région éloignée) et toutes celles qui sont sélectionnés participent (c.-à-d. qu'elles n'ont pas la possibilité de refuser de participer).

Examinons les résultats d'une simulation informatique qui démontrent à quel point le recrutement purement au hasard peut être efficace pour recruter un échantillon représentatif. L'efficacité de la stratégie est évaluée en recrutant de multiples échantillons et en déterminant dans quelle mesure ces derniers reflètent une caractéristique de la population, c'est-à-dire sa moyenne. Dans notre population fictive (simulée à l'aide du logiciel SPSS), 32 000 enseignants ont complété un test de connaissances orthographiques. Les scores au test varient entre 0 et 100 avec une moyenne de 50 pour la population. Dans la simulation, les scores des enseignants sélectionnés sont inclus dans le calcul de la moyenne de l'échantillon.

L'efficacité de la stratégie de recrutement purement au hasard a été examinée en recrutant plusieurs échantillons de trois tailles différentes (graphique 1). La portion supérieure du graphique présente la moyenne de 300 échantillons de 10 enseignants chacun. Malgré leur taille très modeste, les moyennes de ces échantillons sont généralement près de celle de la population. De plus, ils ne sont pas biaisés puisque leur moyenne ne tend pas à sous-estimer ou à surestimer la moyenne de la population. En particulier, il y a autant d'échantillons avec une moyenne inférieure à celle de la population que d'échantillons avec une moyenne supérieure à celle de la population. Néanmoins, ces petits échantillons ne sont pas particulièrement précis. En effet, la moyenne d'une proportion substantielle de ces derniers est inférieure à 45 ou supérieure à 55. Comme l'illustrent les portions médiane et inférieure du graphique 1, la précision s'améliore cependant de façon marquée lorsque la taille des échantillons augmente à 20 ou à 40. Dans ce dernier cas en particulier, les moyennes des échantillons se regroupent étroitement autour de la moyenne de la population. Autrement dit, lorsqu'une stratégie de recrutement purement au hasard est utilisée, même un échantillon de taille relativement modeste est susceptible d'être représentatif et de mener à des conclusions valables.

https://doi.org/10.37571/2022.0204
2022, Vol. 3, N. 2, pp. 103-121



Graphique 1. Recrutement purement au hasard d'échantillons de 10, 20 ou 40 participants[1]

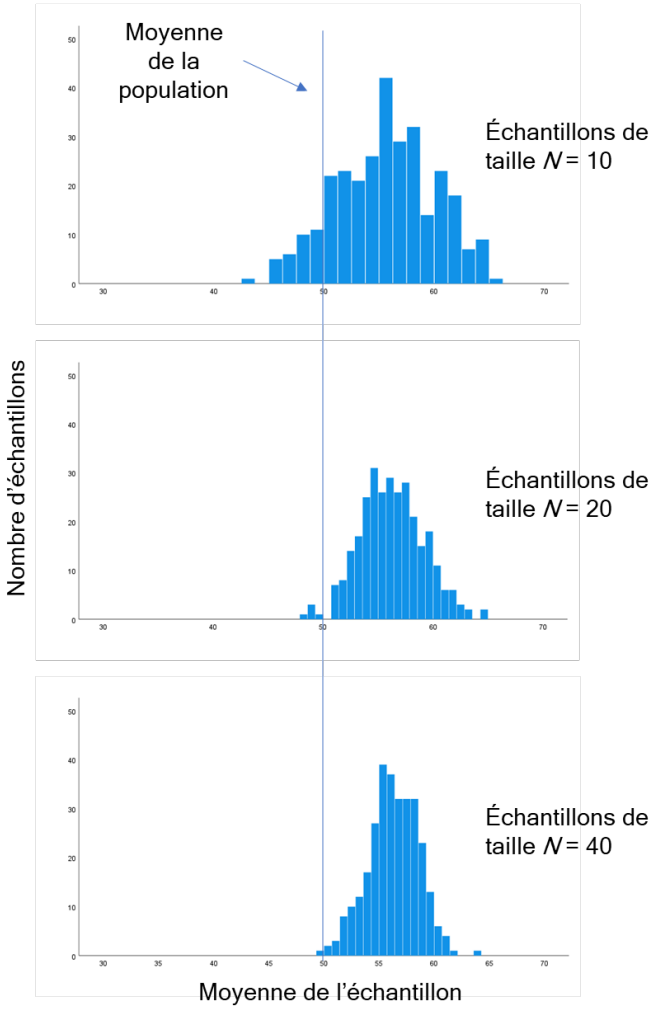
[1] Simulation réalisée à l'aide du logiciel SPSS. Recrutement sans taux de refus au sein d'une population simulée de 32 000 enseignants. Pour chacune des tailles, 300 échantillons sont recrutés au hasard. Les scores varient entre 0 et 100 avec une moyenne populationnelle de 50.

https://doi.org/10.37571/2022.0204
2022, Vol. 3, N. 2, pp. 103-121

Malheureusement, d'un point de vue méthodologique, recruter purement au hasard n'est pas possible en éducation en raison des règles éthiques. En effet, l'éthique de la recherche en sciences sociales dicte en fait que les personnes ont droit de refuser de participer, sans avoir d'ailleurs à expliquer les motifs de leur refus (Conseil de recherches en sciences humaines, Conseil de recherches en sciences naturelles et en génie du Canada, Instituts de recherche en santé du Canada, 2018). Habituellement, un nombre non-négligeable de personnes refusent de participer, souvent pour des raisons qui échappent au chercheur. De tels refus limitent la représentativité s'ils sont le moindrement nombreux *et* s'ils sont en lien avec les objectifs de l'étude.

Pour illustrer la nature du problème, revenons à notre population fictive d'enseignants. Dans cette nouvelle simulation, le recrutement n'est plus purement au hasard. En effet, même si les enseignants sont sélectionnés au hasard, 50% de ceux qui ont obtenu un score faible au test (50 ou moins) refusent de participer. Pour compléter l'échantillon, il est donc nécessaire de contacter davantage d'enseignants, ce qui est susceptible d'entrainer une sur-représentation des enseignants avec un score élevé (ces derniers étant en comparaison plus enclin à participer). Afin de stimuler simplement le plus faible taux de réponse des enseignants avec un score faible, nous avons en fait surreprésenté dans la population les enseignants avec un score élevé. Les portions supérieure, médiane et inférieure du graphique 2 illustrent les conséquences de ces refus pour des tailles d'échantillon de 10, 20 et 40. Les refus introduisent un biais évident puisque les moyennes des échantillons tendent à surestimer la moyenne de la population (qui demeure égale à 50). À noter que plus la taille augmente (parties médiane et inférieure du graphique 2), plus la proportion d'échantillons avec une moyenne approximativement correcte (c.-à-d. d'environ 50) *diminue*. Ainsi, lorsque la stratégie de recrutement introduit un biais, recruter davantage de participants fait en sorte que l'échantillon reflète encore plus clairement le biais.

https://doi.org/10.37571/2022.0204
2022, Vol. 3, N. 2, pp. 103-121



Graphique 2. Recrutement au hasard d'échantillons de 10, 20 ou 40 participants avec un taux de refus de 50% pour les enseignants avec un score sous la moyenne[2]

[2] Note. Simulation réalisée à l'aide du logiciel SPSS. Le taux de refus de 50% pour les enseignants avec un score sous la moyenne (M = 50) est simulée en sur-représentant, dans la population simulée, les enseignants avec un score au-dessus de la moyenne. Pour chacune des tailles, 300 échantillons sont recrutés au hasard. Les scores varient entre 0 et 100 avec une moyenne populationnelle de 50.

https://doi.org/10.37571/2022.0204
2022, Vol. 3, N. 2, pp. 103-121

La stratification et ses limites

Puisque le recrutement purement au hasard est essentiellement impossible en sciences sociales en général et en éducation en particulier, plusieurs méthodologues recommandent de recruter par strates (Zhao, 2021). Une strate est un sous-groupe qui représente une proportion connue de la population. Dans ce cas, le chercheur recrute de manière à que les strates connues soient représentées dans une proportion correcte dans l'échantillon. Par exemple, puisque les filles représentent environ 50% de la population des élèves, le chercheur peut s'assurer que la moitié des participants sont des filles. Les données de recensement de Statistique Canada peuvent être utilisée pour identifier la proportion représentée par d'autres strates (ex. : les familles issues de l'immigration).
Comme le souligne Zhao (2021), le recrutement par strates permet de s'assurer que l'échantillon est *superficiellement* similaire à la population, sans plus. Imaginons qu'un chercheur s'intéresse aux ressources (ex. : aide aux devoirs, soutien professionnel) investies par les familles dans la réussite scolaire des enfants. Comme cet investissement est probablement (imparfaitement) lié au revenu familial, le chercheur consulte les données du recensement pour planifier son recrutement et s'assurer que la diversité de revenu de son échantillon reflètera celle de la population. Même si elle est nécessaire, une telle précaution n'élimine pas la possibilité d'un refus différentiel. À titre d'exemple, l'étude pourrait moins intéresser les familles qui, malgré un bon revenu, investissent relativement peu dans la réussite scolaire des enfants. Bien que l'échantillon soit stratifié par revenu, ces familles sont peu susceptibles de participer et par conséquent d'être correctement représentées.

Considérons un autre exemple d'étude dans lequel une chercheure souhaite établir des normes populationnelles pour un test de mathématiques. Les strates sont ici les niveaux scolaires et la chercheure vise un échantillon de 200 élèves par niveau. Même si tous les parents acceptent que leur enfant participe, ce dernier doit aussi consentir à être évalué. Dans les classes de première à quatrième année, la quasi-totalité des élèves acceptent d'être évalués, ce qui fait en sorte que les différents degrés d'habiletés en lecture sont bien représentés et que les normes sont adéquates. En revanche, en cinquième et en sixième année, le taux de participation est bas chez les élèves faibles en mathématiques. Pour ces deux derniers niveaux, les normes sont artificiellement majorées par la sur-représentation des élèves forts. L'utilisation du test pourrait par conséquent entrainer un surdiagnostic de difficultés en mathématiques chez les élèves de cinquième et sixième année. Puisque la chercheure a réussi à recruter le nombre d'élèves visés pour chaque niveau scolaire, les utilisateurs du test risquent de ne pas soupçonner l'existence du problème, en particulier si

https://doi.org/10.37571/2022.0204
2022, Vol. 3, N. 2, pp. 103-121

la chercheure ne rapporte pas les taux de participation séparément par niveau scolaire dans le manuel du test.

En bref, la stratification est une précaution utile. Cependant, elle ne permet pas d'assurer la représentativité, en raison notamment d'un possible refus différentiel. Ce dernier peut d'ailleurs être indétectable pour le chercheur.

La reproductibilité des résultats

Lorsqu'une étude génère des résultats inattendus et potentiellement importants, les méthodologues recommandent entres autres de reproduire ces résultats auprès d'un nouvel échantillon avant de proposer des modifications aux théories, pratiques, programmes ou politiques (ex. : Duncan, Engel, Claessens et Dowsett, 2014; Mayo-Wilson, Grant et Supplee, 2021). Si plusieurs études reposant sur une méthodologie adéquate génèrent des résultats similaires, les chercheurs, praticiens et décideurs publiques peuvent être relativement confiants que les résultats en question sont fiables. La reproductibilité est encouragée puisqu'aucune étude ne peut être considéré comme infaillible, notamment en raison de ce que Rodgers (1999) appelle les « accidents d'échantillonnage », c'est-à-dire des échantillons qui malgré les efforts des chercheurs s'avèrent complètement non-représentatifs.

Cela dit, la logique de la reproductibilité ne doit pas utilisée pour justifier des pratiques discutables, notamment en matière de recrutement (voir Simmons, Nelson et Simonsohn, 2011). Effectivement, le fait qu'aucune étude ne soit infaillible pourrait encourager les chercheurs à attendre que des collègues reproduisent leurs résultats en utilisant un échantillon plus crédible que le leur. Dans un même ordre d'idées, les chercheurs pourraient considérer que leur étude n'est qu'une parmi d'autre et, par conséquent, que les limites de leur échantillon n'auront pas d'incidence. En d'autres termes, si aucune étude n'est définitive, pourquoi utiliser la méthodologie la plus robuste et tenter de recruter le meilleur échantillon?

Malheureusement, ce qui pourrait être appelé l'attentisme en matière de rigueur scientifique limite les progrès scientifiques en créant une accumulation d'études problématiques qui génère de la confusion (pour un examen des pratiques analytiques ou des devis problématiques, voir Fife et Rodgers, 2021; Slavin et Cheung, 2017). Les résultats utiles et éclairants émanent plutôt d'études bien menées. Ces derniers résultats sont assez crédibles pour faire l'objet de tentatives de reproduction rigoureuses et, lorsqu'ils le sont, d'être effectivement reproduits (pour une démonstration dans le cadre de

https://doi.org/10.37571/2022.0204
2022, Vol. 3, N. 2, pp. 103-121

la recherche en sciences sociales, voir Camerer et al., 2018). Bien qu'il ne s'agisse pas du seul élément, la qualité de l'échantillon contribue inévitablement à la reproductibilité (Fife et Rodgers, 2021; Maxwell, Lau et Howard, 2015; Simmons et al., 2011).

Examinons un exemple de reproductibilité en éducation. Dans l'étude originale, Fuchs, Compton, Fuchs, Paulsen, Bryant et Hamlett (2005) ont testé l'efficacité d'une intervention visant à aider les élèves faibles en mathématiques à surmonter leurs difficultés. L'étude a été menée dans dix écoles primaires. Même si toutes ces écoles étaient situées sur le territoire d'une seule commission scolaire, les chercheurs se sont assurés de représenter une diversité de milieux socio-économiques. Un échantillon d'élèves de taille raisonnable ($N = 139$) a été sélectionné selon des critères spécifiques (des scores faibles à une série de tests). L'intervention en sous-groupe a été offerte par 12 assistants de recherche, pour la plupart étudiants en éducation. Ces derniers ont été formés par la conceptrice de l'intervention (Lynn Fuchs) et étroitement supervisés pendant toute la durée de l'étude (ex. : plusieurs séances d'intervention ont été enregistrées et passées en revue). En comparaison avec leurs pairs assignés au hasard à la condition contrôle (enseignement régulier seulement), les élèves qui ont reçu l'intervention ont fait davantage de progrès à plusieurs évaluations. Étant donné le caractère très contrôlé de cette étude, Gersten, Rolhfus, Clarke, Decker, Wilkins, et Dimino (2015) ont entrepris d'en reproduire les résultats dans une étude à grande échelle menée dans des conditions proches des pratiques scolaires habituelles. Cette dernière étude a été réalisée dans 76 écoles de quatre commissions scolaires auprès d'un échantillon d'élèves sept fois plus grand que celui de Fuchs et collègues. L'intervention a été réalisée par des enseignants remplaçants ou à la retraite formés et accompagnés selon une procédure suffisamment économique pour être normalement utilisables par les écoles. Les résultats de Gersten et collègues ont confirmé que l'intervention était efficace et essentiellement reproduit ceux de Fuchs et collègues (pour un autre exemple récent de reproductibilité, voir Gaspard et al., 2021).

Les tentatives de reproduction comme celles de Gersten et collègues (2015) sont justifiables scientifiquement et finançables lorsque l'étude originale repose sur un devis rigoureux et un échantillon convaincant. En revanche, lorsque la méthodologie d'une étude est problématique, celle-ci n'est pas susceptible de faire l'objet d'une telle tentative sauf si ses résultats sont historiquement importants (ex. : Raudenbush, 1984; Watts, Duncan et Quan, 2018) ou s'ils laissent entrevoir des possibilités d'interventions novatrices (ex. : Kennedy et al., 2017).

https://doi.org/10.37571/2022.0204
2022, Vol. 3, N. 2, pp. 103-121

En règle générale, les chercheurs risquent donc d'être déçus s'ils attendent que d'autres recrutent à leur place des échantillons crédibles pour appuyer leurs résultats. Ils pourraient également être déçu par le fait que leur étude ne soit pas considérée dans les recensions systématiques d'études sur le même sujet, ces recensions appliquant des critères d'inclusion de plus en plus rigoureux (Slavin et Cheung, 2017).

Une stratégie en deux volets

Recruter un échantillon de qualité est nécessaire mais exigeant. En fait, contrairement au problème de la taille, il n'existe pas de solution simple à celui de la représentativité. Nous pensons néanmoins qu'une stratégie en deux volets peut être utile pour éviter les problèmes les plus évidents.

Premièrement, il faut identifier le bassin de recrutement idéal étant donné les objectifs de l'étude en évitant de se tourner simplement vers les participants potentiels les plus accessibles. À titre exemple, si l'étude porte spécifiquement sur l'épuisement des enseignants en poste, il est probablement préférable de ne pas recruter des stagiaires en formation à l'université. Deuxièmement, il faut minimiser le taux de refus au sein du bassin de recrutement idéal. Seul un taux de participation près de 100% protège entièrement contre les biais qui pourraient être introduite par un refus différentiel. Pour maximiser ce taux, les chercheurs peuvent notamment s'assurer que les participants n'ont pas l'impression de perdre leur temps, que les documents qui leur sont destinés sont clairs et concis et que ce sont les chercheurs plutôt que les participants qui, lorsque nécessaire, se déplacent. Le recours à des incitatifs raisonnables peut aussi être utile (CRSH, CRSNG et IRSC, 2018). Si les participants doivent consacrer beaucoup de temps à l'étude, la valeur de cette dernière doit leur être clairement expliquée.

Considérons trois exemples d'études québécoises dans lesquelles la stratégie en deux volets a été déployée par des équipes de chercheurs dont faisaient partie deux des auteurs du présent article. Nous examinons ces études parce que nous savons exactement comment leurs échantillons ont été recrutés et parce que ces derniers ont été considérés comme crédibles par les évaluateurs et les éditeurs de revues internationales.

Dans une étude menée en première année du primaire, Dion, Roux, Landry, Fuchs, Wehby et Dupéré (2011) ont évalué l'efficacité d'activités d'enseignement visant à prévenir les difficultés en lecture. L'étude demandait une implication importante de la part des enseignants. En effet, certains d'entre eux devaient réaliser les activités avec leur groupe à raison de trois périodes de 20 minutes par semaine pendant essentiellement toute l'année. Puisque l'objectif de l'étude était de soutenir la réussite en milieu urbain défavorisé, la

https://doi.org/10.37571/2022.0204
2022, Vol. 3, N. 2, pp. 103-121

participation des écoles du centre de Montréal a été jugée essentielle. Afin d'identifier le bassin de recrutement idéal, le chercheur principal a utilisé les données du recensement sur la défavorisation et l'immigration. Il a ainsi ciblé 44 écoles publiques primaires susceptibles de collectivement refléter la diversité montréalaise notamment en raison de leur dispersion géographique dans les quartiers centraux. Après avoir repéré les coordonnées d'un contact par école, le chercheur a proposé des rencontres en personne dans les écoles. Une rencontre a eu lieu dans 80% des écoles approchées. Lors de la rencontre d'environ 60 minutes, le chercheur a expliqué aux enseignants l'objectif de l'étude, la logique du devis expérimental et leur a présenté des échantillons de matériel d'enseignement. Aucune compensation monétaire n'a été offerte aux enseignants, mais ces derniers pouvaient conserver le matériel à la fin du projet. Le chercheur s'est aussi engagé à distribuer le matériel au meilleur coût à tous les enseignants intéressés si les résultats de l'étude indiquaient qu'il était efficace.

Au total, 79% des enseignants présents lors des rencontres ont décidé de participer. Dans les classes participantes, 83% des parents ont consenti à ce que leur enfant soit évalué. Ces taux de réponse et de consentement se comparent probablement favorablement à ceux observés ailleurs. Il est néanmoins important de reconnaitre que malgré les efforts investis (l'équivalent d'un mois à temps complet), les taux ne sont pas parfaits, ce qui pourrait introduire des biais et limiter la pertinence des recommandations émanant de l'étude. La littérature indique notamment que des activités d'enseignement offertes au groupe ne permettent pas à tous les élèves de réaliser des progrès suffisants (ex. : McMaster, Fuchs, Fuchs et Compton, 2005). Dans des études comme celle de Dion et al., le pourcentage d'élèves participants ne progressant pas est utilisé pour estimer l'ampleur de l'offre des services alternatifs qui devront être offerts par les écoles à l'ensemble de leurs clientèles en difficultés. Pour que le pourcentage observé dans l'étude reflète correctement la situation dans la population scolaire, il faut que le consentement parental ne soit pas influencé par les difficultés de leurs enfants, ce qui apparait improbable (ex. : selon nos conversations avec les enseignants), mais ne peut être démontré (les règles éthiques interdisent la comparaison des élèves avec et sans consentement parental).

Dans leur étude menée au secondaire, Dupéré, Dion, Leventhal, Archambault, Crosnoe et Janosz (2018) se sont plutôt intéressés aux problématiques psychosociales et scolaires qui amènent un adolescent à décrocher de l'école. Étant donné la pression politique et médiatique exercée sur les écoles québécoises pour faire augmenter leur taux de diplomation, ne pas recruter d'écoles publiques avec un fort taux de décrochage aurait sérieusement limité la pertinence locale de l'étude. Les chercheurs ont donc identifié leur

https://doi.org/10.37571/2022.0204
2022, Vol. 3, N. 2, pp. 103-121

bassin de recrutement idéal en consultant les données sur les taux de diplomation par école obtenues par les médias via la loi d'accès à l'information. Comme les faibles taux de diplomation s'observaient à la fois en milieux urbains et ruraux, les chercheurs ont décidé d'approcher également des écoles de ces derniers milieux. Il était prévisible que les écoles urbaines et rurales ciblées seraient réticentes à participer. Afin de minimiser le taux de refus institutionnel, les deux chercheurs principaux ont utilisés leurs réseaux de contacts pour organiser des discussions préliminaires en personne avec les directions. Ces discussions de 60 minutes ont permis d'établir que la méthodologie de l'étude était rigoureuse et originale et que les données obtenues permettraient aux écoles de mieux comprendre les causes probables du décrochage au sein de leurs clientèles d'élèves. Les chercheurs ont aussi donné l'assurance aux directions qu'il serait impossible d'identifier leur école lors de la présentation des résultats.

Au total, 12 des 13 écoles approchées ont accepté de participer. Dans toutes les écoles participantes, les chercheurs principaux ont rencontré l'ensemble des enseignants en début d'année scolaire pour leur expliquer le projet. Finalement, les chercheurs ont fait valoir qu'il était probable que plusieurs adolescents décrocheurs provenaient de familles en difficulté. Par conséquent, ils ont obtenu l'autorisation du comité éthique de se soustraire à l'exigence d'obtenir le consentement parental. Cette dérogation, l'appuie enthousiaste des équipes-écoles et le tirage de certificats cadeaux ont permis d'obtenir un taux de participation de 98% aux questionnaires de dépistage complété par les élèves en début d'année. Parmi les élèves approchés au cours des mois suivants, 70% ont accepté de participer à une entrevue structurée (un incitatif financier leur a été offert), ce qui représente un pourcentage élevé considérant que le tiers de ces élèves venaient tout juste de décrocher de l'école. Collectivement, les deux chercheurs principaux ont consacré trois mois de travail à plein pour recruter les 12 écoles et les 545 élèves interviewés.

Les taux de participation obtenus par Dupéré et al. (2018) se comparent favorablement à ceux rapportés dans la littérature scientifique. Encore une fois cependant, le fait que ces taux ne soient pas parfaits a pu introduire des biais à différents niveaux. Les chercheurs ont notamment l'impression que la direction de l'école qui a refusé de participer n'était pas à l'aise avec sa gestion du décrochage, une dynamique institutionnelle probablement sous représentée dans l'échantillon. De plus, les discussions avec les équipes écoles ainsi que les brefs contacts téléphoniques avec les adolescents qui ont refusés d'être interviewés suggèrent que ces derniers vivaient des situations particulièrement difficiles. Même si de nombreuses problématiques aigues et embarrassantes (ex. : de victimisation intense) ont été rapportées en entrevue, d'autres ont pu être occultées par les refus de participation. Il

https://doi.org/10.37571/2022.0204
2022, Vol. 3, N. 2, pp. 103-121

est donc possible que les données de Dupéré et al. sous-estiment le rôle des problèmes de gestion scolaire et des difficultés psychosociales dans le phénomène du décrochage.

Dans leur étude longitudinale, Plante, Lecours, Lapointe, Chaffee et Fréchette-Simard (2022) se sont intéressés aux liens entre la réussite au primaire et l'anxiété lié aux évaluations au moment de la transition au secondaire. Les chercheurs ont sélectionné un bassin d'écoles de milieux ruraux et de banlieue avec deux objectifs en tête : inclure la plus grande diversité possible (ex. : en termes de réussite) tout en minimisant l'attrition au moment de la transition primaire-secondaire. Ils ont donc recruté dans des secteurs socio-économiquement diversifiés où les élèves ne se dispersaient pas dans plusieurs écoles publiques ou privées au secondaire. Lors de conférences téléphoniques, les chercheurs ont expliqué aux directions le déroulement et les retombées potentielles de l'étude ainsi que l'importance de minimiser le taux d'attrition. Les 28 écoles primaires et 11 écoles secondaires contactés ont accepté de participer. À la suggestion d'un représentant de la direction d'un centre de service scolaire, les chercheurs ont également mis en place une stratégie pour maximiser le taux de retour des formulaires de consentement parental. La classe a reçu un budget équivalent à 5$ par formulaire remis à l'enseignant indépendamment de la décision du parent. Cette stratégie a permis d'obtenir un taux de participation de 95%.

Le taux obtenu par Plante et al. (2022) pourrait être jusqu'à 20% supérieur à celui observé généralement dans les études du genre (Tigges, 2003). Néanmoins, des discussions informelles avec les enseignants suggèrent que les quelques parents qui n'ont pas remis le formulaire de consentement ne répondaient généralement pas aux demandes administratives de l'école. Il est donc probable que les élèves de familles très désorganisées ou désengagées soient sous-représentés dans l'étude.

En bref, toute personne est libre de refuser de participer. Les chercheurs peuvent néanmoins identifier les participants éventuels qui sont les plus importants pour leurs études (volet 1) et, en étant sensible aux besoins et aux préoccupations de ces personnes, leur donner envie de participer (volet 2).

Conclusion

L'objectif du présent article était d'identifier les caractéristiques d'échantillons de qualité et de formaliser une stratégie réaliste pour recruter de tels échantillons. Nous nous sommes attardés en particulier à l'importance et à la difficulté de recruter des échantillons représentatifs. À l'issue de ce que nous avons abordé dans cet article, nous pensons qu'il

https://doi.org/10.37571/2022.0204
2022, Vol. 3, N. 2, pp. 103-121

est opportun de réfléchir à la représentativité des échantillons. Considérons un exemple historique pour établir le contraste avec la situation actuelle. Il y a 40 ans, Beck, Perfetti et McKeown (1982) publiaient une étude sur l'efficacité de l'enseignement explicite du vocabulaire. Dans leur étude, les activités d'enseignement étaient implantées par une seule enseignante avec son groupe. Deux autres classes poursuivaient leurs activités régulières. En plus d'être de taille modeste, l'échantillon n'était décrit que sommairement dans l'article. L'étude a néanmoins été jugée, à l'époque, comme suffisamment de qualité pour être publiée dans une revue scientifique de premier plan.

Aujourd'hui, une telle étude serait probablement considérée comme un pilote publiable dans une revue mineure, entre autres en raison de ce qui seraient maintenant considérés comme des problèmes de devis et d'analyses évidents (Slavin et Cheung, 2017; What Works Clearinghouse, 2021) et de l'analyse statistique (Raudenbush et Bryk, 2001; Roberts, Scammacca et Roberts; 2018). L'échantillon de l'étude serait également considéré comme de très petite taille, même pour un pilote. De plus, même en l'absence de critère formel, les évaluateurs douteraient probablement de sa représentativité. Cette sensibilité accrue à la représentativité se manifeste notamment dans l'utilisation des diagrammes de flux (ex. : Gaspard et al., 2021; Michaud, Dion, Barrette, Dupéré et Toste, 2017; Montero-Marin et al., 2021). De tels diagrammes servent à décrire relativement en détails le recrutement et permettent d'estimer la représentativité de l'échantillon.

En conclusion, le présent article s'est attardé à la représentativité plutôt qu'à la taille des échantillons puisqu'une vaste littérature méthodologique traite déjà de ce deuxième aspect (ex. : Luo et al., 2021; Zhang, Spybrook et Unlu, 2020). La représentativité, qui a longtemps été négligée, doit être à notre avis ramenée au cœur de la réflexion. La centration des méthodologues sur taille pourrait avoir amené certains chercheurs à considérer que cette caractéristique est la seule qui compte. Il est urgent de reconnaitre que ce n'est pas le cas. L'arrivée de l'internet, des médias sociaux et des logiciels de sondage en ligne a facilité le recrutement de grands échantillons de toutes évidences peu représentatifs. Le fait que le recrutement à distance soit devenu une option attrayante, notamment pour les chercheurs en didactique, ne doit pas nous faire perdre de vue l'exigence de rigueur.

https://doi.org/10.37571/2022.0204
2022, Vol. 3, N. 2, pp. 103-121

Références

Aruguete, M. S., Huynh, H., Browne, B. L., Jurs, B., Flint, E. et McCutcheon, L. E. (2019). How serious is the 'carelessness' problem on Mechanical Turk? *International Journal of Social Research Methodology, 22*, 441-449.

Beck, I. L., Perfetti, C. A. et McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology, 74*(4), 506-521. https://doi.org/10.1037/0022-0663.74.4.506

Berkeley Statistics (2021). *Glossary of statistical terms*. https://www.stat.berkeley.edu/~stark/SticiGui/Text/gloss.htm

Buhrmester, M. D., Talaifar, S., et Gosling, S. D. (2018). An evaluation of Amazon's Mechanical Turk, its rapid rise, and its effective use. *Perspectives on Psychological Science, 13*, 149-154.

Camerer, C. F., Dreber, A., Holzmeister, F. Ho, T.-H., Huber, J., Johannesson, M., Kirchler, Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T. (…) & Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behavior*, 2, 637–644. https://doi-org.proxy.bibliotheques.uqam.ca/10.1038/s41562-018-0399-z

Conseil de recherches en sciences humaines [CRSH], Conseil de recherches en sciences naturelles et en génie du Canada [CRSNG], Instituts de recherche en santé du Canada [IRSC] (2018). *Énoncé de politique des trois Conseils : Éthique de la recherche avec des êtres humains*. Auteurs.

Dion, E., Roux, C., Landry, D., Fuchs, D., Wehby, J. et Dupéré, V. (2011). Improving attention and preventing reading difficulties among low-income first-graders. *Prevention Science*, *12*(1), 70-79. https://doi.org/10.1007/s11121-010-0182-5

Duncan, G. J., Engel, M., Claessens, A., et Dowsett, C. J. (2014). Replication and robustness in developmental research. *Developmental Psychology, 50*(11), 2417-2425. https://doi.org/10.1037/a0037996

Dupéré, V., Dion, E., Leventhal, T., Archambault, I., Crosnoe, R. et Janosz, M. (2018). High school dropout in proximal context: The triggering role of stressful life events. *Child Development*, *89*(2), e107-e122. https://doi.org/10.1111/cdev.12792

Fife, D. A. et Rodgers, J. L. (2021, November 15). Understanding the exploratory/confirmatory data analysis continuum: Moving beyond the "replication crisis". *American Psychologist*. Prépublication en ligne. http://dx.doi.org/10.1037/amp0000886

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., et Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97*(3), 493–513. https://doi-org.proxy.bibliotheques.uqam.ca/10.1037/0022-0663.97.3.493

Gaspard, H., Parrisius, C., Piesch, H., Kleinhansl, M., Wille, E., Nagengast, B., Trautwein, U. et Hulleman, C. S. (2021). The potential of relevance interventions for scaling up: A cluster-randomized trial testing the effectiveness of a relevance intervention in math classrooms. *Journal of Educational Psychology, 113*(8), 1507-1528. https://doi.org/10.1037/edu0000663

Gersten R., Rolfhus E., Clarke B., Decker L. E., Wilkins C. et Dimino J. (2015). Intervention for first graders with limited number knowledge: Large-scale replication of a randomized controlled trial. *American Educational Research Journal, 52*(3), 516-546. https://doi.org/10.3102/0002831214565787

Jensen-Doss, A., Patel, Z. S., Casline, E., Mora Ringle, V. A. et Timpano, K. R. (2021). Using Mechanical Turk to study parents and children: An examination of data quality and representativeness. *Journal of Clinical Child & Adolescent Psychology*, 1-15. Prépublication en ligne.

Kennedy, M. J., Hirsch, S. E., Rodgers, W. J., Bruce, A. et Lloyd, J. W. (2017). Supporting high school teachers' implementation of evidence-based classroom management practices. *Teaching and Teacher Education, 63*, 47-57. https://doi.org/10.1016/j.tate.2016.12.009

Kennedy, R., Clifford, S., Burleigh, T., Jewell, R. et Waggoner, P. D. (2018). The shape of and solutions to the MTurk quality crisis. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3272468

Luo, W., Li, H., Baek, E., Chen, S., Lam, K. H., et Semma, B. (2021). Reporting practice in multilevel modeling: A revisit after 10 years. *Review of Educational Research, 91*(3), 311-355. https://doi.org/10.3102/0034654321991229

Maxwell, S. E., Lau, M. Y. et Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist, 70*(6), 487-498. https://doi.org/10.1037/a0039400

Mayo-Wilson, E., Grant, S. et Supplee, L. H. (2021). Clearinghouse standards of evidence on the transparency, openness, and reproducibility of intervention evaluations. *Prevention Science*. Prépublication en ligne. https://doi.org/10.1007/s11121-021-01284-x

McMaster, K. L., Fuchs, D., Fuchs, L. S., et Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention

methods. *Exceptional Children, 71*(4), 445-463. https://doi.org/10.1177/001440290507100404

Michaud, M., Dion, E., Barrette, A., Dupéré, V. et Toste, J. (2017). Does knowing what a word means influence how easily its decoding is learned? *Reading & Writing Quarterly, 33*(1), 82-96. http://dx.doi.org/10.1080/10573569.2015.1092003

Montero-Marin, J., Taylor, L., Crane, C., Greenberg, M. T., Ford, T. J., Williams, J. M. G., García-Campayo, J., Sonley, A., Lord, L., Dalgleish, T., Blakemore, S.-J., MYRIAD team et Kuyken, W. (2021). Teachers "finding peace in a frantic world": An experimental study of self-taught and instructor-led mindfulness program formats on acceptability, effectiveness, and mechanisms. *Journal of Educational Psychology, 113*(8), 1689-1708. https://doi-org.proxy.bibliotheques.uqam.ca/10.1037/edu0000542

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., et Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health, 42*(5), 533-544. https://doi.org/10.1007/s10488-013-0528-y

Plante, I., Lecours, V., Lapointe, R., Chaffee, K. E., et Fréchette-Simard, C. (2022). Relations between prior school performance and later test anxiety during the transition to secondary school. *British Journal of Educational Psychology*. Prépublication en ligne. https://doi.org/10.1111/bjep.12488

Réseau d'information pour la réussite éducative (2021). *Lignes* directrices *pour la soumission d'articles*. http://rire.ctreq.qc.ca/lignes-directrices-pour-la-soumission-darticles/

Raudenbush, S. W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from 18 experiments. *Journal of Educational Psychology, 76*(1), 85-97. https://doi.org/10.1037/0022-0663.76.1.85

Raudenbush, S. W. et Bryk, A. S. (2001). Hierarchical linear models: Applications and data analysis methods (2e édition). Sage.

Roberts, G., Scammacca, N. et Roberts, G. J. (2018). Causal mediation in educational intervention studies. *Behavioral disorders, 43*(4), 457–465. https://doi.org/10.1177/0198742917749560

Robinson, O. C. (2014). Sampling in interview-based qualitative research: A theoretical and practical guide. *Qualitative Research in Psychology, 11*(1), 25-41. https://doi.org/10.1080/14780887.2013.801543

Simmons, J. P., Nelson, L. D. et Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as

significant. *Psychological Science, 22*(11), 1359-1366. https://doi.org/10.1177/0956797611417632

Slavin, R. E. et Cheung, A. C. K. (2017). Lessons learned from large-scale randomized experiments. *Journal of Education for Students Placed at Risk, 22*(4), 253-259. https://doi.org/10.1080/10824669.2017.1360774

Tigges, B. B. (2003). Parental consent and adolescent risk behavior research. *Journal of Nursing Scholarship, 35*(3), 283-289. https://doi.org/10.1111/j.1547-5069.2003.00283

Watts, T. W., Duncan, G. J., et Quan, H. (2018). Revisiting the marshmallow test: A conceptual replication investigating links between early delay of gratification and later outcomes. *Psychological Science, 29*(7), 1159–1177. https://doi-org.proxy.bibliotheques.uqam.ca/10.1177/0956797618761661

What Works Clearinghouse (2021). *Standards Handbook, version 4.1*. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf

Zhang, Q., Spybrook, J. et Unlu, F. (2020). Examining design and statistical power for planning cluster randomized trials aimed at improving student science achievement and science teacher outcomes. *AERA Open, 6*(3), 1-12. https://doi.org/10.1177%2F2332858420939526

Zhao, K. (2021). Sample representation in the social sciences. *Synthese, 198*, 9097–9115 (2021). https://doi.org/10.1007/s11229-020-02621-3