

Étude sur la puissance statistique des devis de recherche en éducation

The statistical power of research designs in education

Estudio sobre la potencia estadística de los diseños de investigación en educación

Éric Yergeau

Volume 35, Number 2, 2009

Les manuels scolaires : réformes curriculaires, développement professionnel et apprentissages des élèves

URI: <https://id.erudit.org/iderudit/038735ar>

DOI: <https://doi.org/10.7202/038735ar>

[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

Yergeau, É. (2009). Étude sur la puissance statistique des devis de recherche en éducation. *Revue des sciences de l'éducation*, 35(2), 199–221.
<https://doi.org/10.7202/038735ar>

Article abstract

This article examines the mean statistical power of recent articles published in French Canadian scientific journals dedicated to research in education. 77 studies published in 4 different journals between 2004 and 2006 were analyzed with G*Power to establish their potential to detect small, medium, and large effects. Overall, results show that little progress has been made since the founding work of Cohen (1962) regarding statistical power in the field of educational research in the francophone scientific community. The author gives concrete actions for researchers to improve the statistical power of future work.

Étude sur la puissance statistique des devis de recherche en éducation



Éric Vergeau, professeur
Université de Sherbrooke

RÉSUMÉ • Cet article propose de faire le point sur la puissance statistique moyenne des devis de recherche en éducation publiés récemment au Canada francophone. Soixante-dix-sept articles publiés entre 2004 et 2006, à l'intérieur de quatre revues scientifiques canadiennes francophones, ont été analysés avec le logiciel *G*Power* pour déterminer leur capacité à détecter des effets de petite, moyenne et grande taille. Les résultats montrent peu de progrès réels de la recherche en éducation au Canada français en matière de puissance statistique depuis les travaux de Cohen (1962). Des pistes concrètes pour améliorer la puissance statistique des devis sont suggérées.

MOTS CLÉS • puissance statistique, recherche quantitative, devis de recherche, statistiques inférentielles, validité interne.

1. Introduction et problématique

Les développements technologiques de la dernière décennie et plus spécifiquement les avancées en matière de traitement informatisé des données ont permis, entre autres, aux utilisateurs des méthodes quantitatives de passer des analyses descriptives et bivariées aux stratégies d'analyse multivariées. Ces progrès importants ont ouvert la porte à l'exploration de questions de recherche de plus en plus élaborées qui, à leur tour, font appel à des techniques statistiques beaucoup plus complexes. Nonobstant cette évolution fulgurante de la capacité d'analyse, les techniques statistiques inférentielles reposent toujours sur les mêmes préceptes théoriques que Fisher (1956) a proposés en regard de la démarche scientifique, à savoir la vérification de l'hypothèse nulle. Cette conception de la démarche scientifique stipule que, devant l'incapacité de la science de prouver empiriquement, hors de tout doute, l'existence d'un phénomène dans la réalité, l'alternative rationnelle consiste à évaluer l'erreur commise en avançant que le phénomène est présent

(hypothèse de recherche ou H_1), si celui-ci est en réalité inexistant (hypothèse nulle ou H_0). Les propos de Fisher (1956) ont trouvé écho, chez les lecteurs, car le calcul statistique de la probabilité d'erreur associée au rejet de l'hypothèse nulle comme fondement de la démarche scientifique fait maintenant consensus dans le monde de la recherche en général. À cet égard, les résultats présentés dans les publications scientifiques en éducation et dans beaucoup d'autres domaines montrent bien l'importance accordée au degré de certitude (basé sur le seuil de signification) avec lequel on peut rejeter l'hypothèse nulle et mettre de l'avant l'hypothèse alternative de recherche. À ce sujet, les critiques de Neyman et Pearson au sujet de l'inférence statistique, bien résumées dans les travaux de Marks (2003) et plus récemment de Bourque, Gingras et Larose (2009), ont provoqué un débat encore très actuel dans le monde scientifique. En bref, ces auteurs ont avancé que la démarche empirique est en fait un processus décisionnel, et que les conclusions d'une étude ne doivent pas reposer uniquement sur le test d'hypothèse, mais également sur le contexte expérimental. Selon eux, les enjeux entourant la validité du rejet de l'hypothèse nulle devrait occuper une place prépondérante dans l'analyse des résultats de recherche. Sans entrer dans les détails de la controverse, on remarque que l'utilisation contemporaine des techniques statistiques incorpore de manière concomitante les cadres conceptuels de Fisher (1956) et de Neyman-Pearson (Neyman, 1942). En ce sens, on aurait tendance à croire que la vérification des conditions optimales de rejet de l'hypothèse nulle est effectuée de manière routinière par les chercheurs. Malheureusement, selon Rossi (1990) et Cohen (1990, 1994), tel ne semble pas être le cas. Selon ces auteurs, beaucoup de chercheurs ne prêtent que peu d'attention à la puissance statistique de leur devis. Dans la foulée des travaux fondateurs de Cohen (1962) sur le sujet, plusieurs chercheurs ont tenté de sensibiliser la communauté scientifique à l'importance centrale du concept de puissance statistique dans l'élaboration des devis de recherche et dans le test d'hypothèses, et ce, avec plus ou moins de succès (Maddock et Rossi, 2001 ; Rossi, 1990 ; Sedlmeier et Gigerenzer, 1989). D'où notre question de recherche qui est la suivante : *Quel est l'état actuel des devis de recherche en éducation en matière de puissance statistique ?*

2. Contexte théorique

2.1 Puissance statistique

Classiquement, la puissance statistique se définit comme la probabilité, à l'intérieur d'un devis de recherche, de rejeter l'hypothèse nulle lorsque celle-ci est réellement fautive dans la population (Cohen, 1988). Dans ce contexte, un devis de recherche avec peu de puissance statistique aura une probabilité beaucoup plus faible de conduire au rejet de l'hypothèse nulle qu'un devis ayant une puissance plus grande. À titre d'exemple, un devis ayant une puissance statistique de 0,50 a, dès le départ, une chance sur deux de rejeter l'hypothèse nulle lorsque celle-ci est effectivement fautive dans la population. Cela veut dire que, pour une étude utilisant toujours le

même devis et répliquée auprès de 100 échantillons différents, 50 % de ces répliqués rapporteront des résultats *significatifs*, tandis que les autres ne rejeteront pas H_0 bien que celle-ci soit fautive. Ici, on comprend bien qu'un champ d'étude où les devis présentent en moyenne une faible puissance statistique puisse générer des résultats équivoques et susciter la controverse.

Depuis les quatre dernières décennies, les quelques travaux de méta-analyse portant sur les études publiées dans divers domaines de la recherche psychosociale montrent invariablement que les devis possèdent généralement une faible puissance statistique et, par conséquent, que les chercheurs semblent se préoccuper bien peu de cette dimension (Brewer, 1972; Cohen, 1962, 1992; Maddock et Rossi, 2001; Orme et Tolman, 1986; Rossi, 1990). Reprenant les travaux de Cohen (1962), Sedlmeier et Gigerenzer (1989) ont montré que, 25 ans après les conclusions de Cohen, le niveau moyen de la puissance statistique des recherches n'avait pratiquement pas évolué. Pourtant, les conséquences d'une faible puissance statistique sur la validité des résultats de recherche sont bien tangibles et devraient attirer l'attention des chercheurs. En effet, comme nous l'avons illustré, une recherche n'ayant pas suffisamment de puissance statistique peut permettre de conclure à l'absence d'impact d'une intervention éducative, alors que cet effet existe pourtant bel et bien. Une telle conclusion pourrait conduire les décideurs à abolir une intervention efficace parce que le protocole d'évaluation n'a pas tenu compte de la puissance statistique nécessaire pour détecter son impact. C'est ce qu'ont constaté empiriquement Kazdin et Bass (1989). Les résultats qu'ils présentent montrent que le manque de puissance statistique est à l'origine des résultats équivoques concernant l'évaluation de l'efficacité des psychothérapies. À la lumière de ce qui précède, Cohen (1962, 1988, 1990) recommande fortement de calculer systématiquement la puissance statistique des devis avant de procéder à la cueillette des données ou d'en publier les résultats.

2.2 Déterminants de la puissance statistique

Cohen (1988) explique que la puissance statistique d'un devis est fonction de trois déterminants interdépendants. En fait, les quatre éléments suivants : la puissance statistique, le degré de signification, la taille de l'échantillon et la taille de l'effet sont en relation, de telle sorte que lorsque trois d'entre eux sont connus, le dernier peut être déduit ; ainsi, la variation de la valeur d'un élément fait fluctuer celle des trois autres.

2.2.1 Définition de la puissance statistique

Nous pouvons maintenant préciser la définition de la puissance statistique d'un test en spécifiant qu'elle représente *la probabilité de rejeter correctement l'hypothèse nulle en fonction d'un seuil de signification, d'une taille d'échantillon et d'une taille d'effet prédéterminés*. Conséquemment, l'incapacité de rejeter l'hypothèse nulle alors qu'elle est fautive produit l'erreur dite *de Type II*. La probabilité

d'apparition de cette erreur est représentée par le seuil *beta* (β). La puissance statistique correspond donc mathématiquement à $1 - \beta$, soit la probabilité de rejeter H_0 lorsqu'elle est fausse.

2.2.2 Seuil de signification

Un autre paramètre à considérer est le degré de signification désiré par le chercheur (seuil *alpha*, α) ou encore la probabilité maximale de rejeter l'hypothèse nulle alors que cette dernière est vraie dans la population (erreur dite *de Type I*). Traditionnellement, ce seuil est fixé à 5% mais il est possible que certaines situations de recherche demandent un seuil plus restrictif ou plus souple. À cet égard, notons qu'il est possible d'augmenter la puissance statistique d'un test en assouplissant le seuil de signification, en le faisant passer par exemple de 0,05 à 0,10.

2.2.3 Taille de l'échantillon

La taille de l'échantillon a un effet direct sur la puissance statistique d'une étude. Dans la mesure où l'augmentation de la taille d'un échantillon réduit l'erreur-type de l'estimation d'un paramètre, elle augmente, du même coup, la puissance statistique d'un devis. En ce sens, la détermination de la taille de l'échantillon d'une étude est une étape cruciale dans la construction d'un devis de recherche, car cet élément est celui qui a le plus d'impact sur le calcul de la puissance statistique.

2.2.4 Taille de l'effet

La taille de l'effet représente l'écart entre l'hypothèse nulle et l'hypothèse de recherche ou, en d'autres mots, l'ampleur avec laquelle l'hypothèse nulle est fautive dans la population. Ce paramètre est le plus difficile à circonscrire car, d'une part, on a bien peu de données précises sur la magnitude des phénomènes humains étudiés en général et spécifiquement dans le domaine de l'éducation et, d'autre part, chaque test statistique possède ses balises de taille d'effet. Pour bien comprendre ce concept, Cohen (1992) explique qu'une taille d'effet de 0 correspond en fait à l'hypothèse nulle. Par exemple, l'hypothèse nulle du test de corrélation de Pearson est que $r = 0$ et pour les tests de différence de moyennes (le test-*t* de Student par exemple), que ces différences entre les groupes égalent 0. Tout écart de l'hypothèse nulle représente un effet dont la taille est quantifiable par un indice positif qui se calcule différemment selon le test employé. Dans le but de simplifier la compréhension et l'utilisation de la taille de l'effet, Cohen (1962, 1988) a établi une convention qui ramène le calcul pour chaque test à des dimensions communes : l'effet de petite, de moyenne et de grande taille. En termes opérationnels, l'effet de taille moyenne se veut un effet visible à l'œil nu pour un observateur attentif. L'effet de petite taille décrit un effet plus subtil que l'effet de taille moyenne sans pour autant être cliniquement insignifiant, tandis que l'effet de grande taille est à la même distance au-dessus de l'effet de taille moyenne que l'effet de petite taille lui est inférieur. Bien qu'arbitraire, cette convention fait tout de même office de réfère-

rence dans la grande majorité des textes portant sur la puissance statistique et sur les tests d'hypothèses.

En résumé, la puissance statistique d'un test est fonction du seuil de signification, de la taille de l'échantillon et de la taille de l'effet du phénomène à l'étude.

2.3 Analyse de la puissance statistique *a priori*

L'application la plus pertinente des concepts expliqués précédemment est certainement l'analyse de la puissance statistique d'un devis de recherche lors de sa conception à l'étape de la planification d'une étude scientifique. Par cette analyse, le chercheur vise à établir la combinaison optimale entre les quatre composantes que sont la puissance statistique, le seuil de signification, la taille de l'échantillon et la taille de l'effet à détecter. Habituellement, cette analyse sert, en priorité, à calculer la taille optimale de l'échantillon nécessaire pour le maintien d'une puissance statistique adéquate en regard du plan d'analyses statistiques prévu ultérieurement par le chercheur. Pour ce faire, on fixe les trois autres déterminants de la manière suivante: le seuil de signification est normalement réglé à 5 % et une puissance statistique autour de 0,80 est jugée suffisante par Cohen (1988). Il reste donc au chercheur à établir la taille de l'effet qu'il pense devoir être en mesure de détecter pour répondre à ses questions de recherche.

Cette étape est la plus ardue car, comme nous l'avons mentionné auparavant, il existe peu d'informations claires sur la magnitude des effets produits dans la population par les divers phénomènes étudiés dans le domaine de l'éducation. À ce chapitre, Orme et Tolman (1986) proposent trois stratégies pour juger de la taille de l'effet d'un phénomène. Premièrement, l'effet peut être estimé à l'aide des travaux de recherche antérieurs portant sur le même thème. Une revue des écrits sur un thème spécifique peut donner des indications générales relatives à l'effet qui est observé dans des recherches comparables. Deuxièmement, le chercheur peut s'appuyer sur des éléments théoriques pour proposer une taille d'effet plausible du phénomène, ou encore il peut l'inférer à partir de considérations purement pratiques. En effet, dans le contexte de certaines études, seuls les effets qui sont suffisamment importants pour être non équivoques peuvent intéresser le chercheur et, ainsi, la capacité à détecter des effets subtils devient, en pratique, inutile. La troisième alternative proposée par ces auteurs est de classer l'effet du phénomène à l'intérieur des balises avancées par Cohen (1988), soit celles de petite, moyenne et grande taille. Ces conventions ont démontré leur utilité et sont employées couramment et opérationnellement dans les écrits scientifiques.

Par la suite, lorsque ces trois déterminants sont fixés, le chercheur peut se référer aux tableaux de Cohen (1988) ou encore à des logiciels spécialisés (ex.: *G*Power*, développé par Faul et Erdfelder, 1992) qui permettent d'évaluer, de manière approximative, le nombre de sujets nécessaire pour parvenir à une puissance statistique adéquate. En effectuant cette analyse, le chercheur évite de recruter un échantillon trop imposant lorsque l'effet à détecter est grand (donc il réalise une économie

substantielle) et, inversement, il peut ajuster à la hausse la taille de l'échantillon lorsque l'effet d'intérêt est plus subtil. À la lumière de ces constats, il devient évident que les efforts consacrés à l'analyse préalable de la puissance statistique et de ses déterminants consolident la validité interne et externe d'un devis de recherche. Cela dit, il est également possible d'analyser *a posteriori* la puissance statistique d'un devis.

2.4 Analyse de la puissance statistique *a posteriori*

Une fois que les résultats d'une étude sont disponibles, il est aussi possible d'en calculer la puissance statistique, et ce, toujours par le truchement des relations existant entre les déterminants de la puissance statistique. Pour chaque test statistique réalisé, l'examen des résultats permet d'identifier la taille de l'échantillon utilisé ainsi que le seuil de signification retenu. Cependant, comme il est très difficile de déterminer clairement la taille de l'effet qui est théoriquement à détecter par une étude spécifique, il s'avère plus pertinent d'estimer la taille de celle-ci en se servant des catégories prédéterminées de petite, moyenne et de grande taille d'effet (Cohen, 1988). On peut alors calculer la puissance d'un devis à détecter un effet de petite, moyenne ou grande taille. Ce type d'analyse *a posteriori* est surtout employé dans des travaux de recension pour estimer la puissance statistique moyenne des devis de recherche dans un champ spécifique.

2.5 Recension des écrits

À cet égard, dans le tableau 1, est dressée une liste chronologique des principales recensions visant à estimer la puissance statistique moyenne des devis d'études publiées dans le vaste domaine de la recherche en sciences humaines depuis plus de 40 ans. Dans la foulée de l'effort de Cohen (1962), la plupart des travaux subséquents sur la puissance statistique ont reproduit sa démarche. En effet, ces recensions portent généralement sur l'ensemble des articles publiés par un ou quelques journaux scientifiques pendant une année, et qui présentent des résultats basés sur des analyses statistiques inférentielles. Ce respect de la démarche de Cohen (1962) à l'intérieur des autres travaux de recension permet d'examiner, en termes de progrès ou de recul, l'évolution de la puissance statistique dans les recherches en sciences humaines et sociales au fil des décennies. Dans l'ensemble, on remarque que la puissance statistique moyenne des recherches recensées puis compilées par toutes ces études n'a pas beaucoup augmenté depuis 1962, date de la publication de l'article de Cohen, et ce, toutes disciplines confondues.

Les résultats de notre recherche bibliographique ont permis de repérer une seule étude ayant eu comme sujet d'analyse la puissance statistique des études scientifiques dans le domaine de l'éducation, soit l'article de Brewer (1972). Ce texte traite des articles publiés dans une seule revue américaine entre les années 1959 et 1971. Les résultats de cette étude indiquent que les recherches menées pendant cette période et publiées dans ce journal avaient en moyenne une faible puissance statistique. La puissance moyenne pour détecter des effets de petite taille variait entre

Tableau 1
Résultats obtenus par les recensions portant sur la puissance statistique

Auteur (année)	Domaine	Période	Nombre d'articles	Taille de l'effet (médiane)		
				Petite	Moyenne	Grande
Cohen (1962)	Psychologie	1960	70	0,18 (0,17)	0,48 (0,46)	0,83 (0,89)
Brewer (1972)	Éducation	1969-71	47	0,14 (N/D)	0,58 (N/D)	0,78 (N/D)
Haase (1974)	Counselling	1968-71	60	0,19 (0,10)	0,46 (0,37)	0,72 (0,74)
Orme et Tolman (1986)	Travail social	1976-85	64	0,20 (0,14)	0,68 (0,78)	0,88 (0,99)
Magid et collab. (1987)	Management	1982-84	44	0,23 (0,13)	0,59 (0,53)	0,83 (0,88)
Sedlmeier et Gigerenzer (1989)	Psychologie	1984	54	0,21 (0,14)	0,50 (0,44)	0,84 (0,90)
Rossi (1990)	Psychologie	1982	312	0,17 (N/D)	0,57 (N/D)	0,83 (N/D)
Maddock et Rossi (2001)	Santé et psychologie	1997	216	0,20 (N/D)	0,68 (N/D)	0,88 (N/D)

0,11 et 0,14; entre 0,45 et 0,71 pour les effets de taille moyenne, et entre 0,69 et 0,85 pour les effets de grande taille. Mis à part ce texte, il a été impossible de trouver des travaux récents portant sur l'état de la puissance statistique des devis de recherche dans le domaine de l'éducation, et encore moins dans la sphère francophone d'activités scientifiques. Sur le plan de la puissance, on peut donc se questionner sur la qualité globale des recherches en éducation selon cet aspect spécifique. Est-ce que les recherches produites par les chercheurs du Canada francophone en éducation sont comparables à celles des autres domaines scientifiques en sciences sociales en matière de puissance statistique? À la lumière de l'importance qu'une puissance statistique suffisante apporte à la validité des conclusions d'une étude, il nous apparaît pertinent, voire essentiel, pour le domaine de la recherche en éducation, d'effectuer un travail rigoureux sur cette question.

L'objectif principal de cet article est donc de brosser le portrait de la puissance statistique des devis de recherche sur lesquels s'appuient les résultats des études francophones canadiennes récemment publiées dans le domaine de l'éducation.

3. Méthodologie

3.1 Sélection des articles scientifiques

Dans cette étude, nous examinons un échantillon de 77 articles scientifiques, identifiés par un astérisque dans la section *Références*, et publiés entre 2004 et 2006 dans une des quatre revues canadiennes francophones suivantes: la *Revue des sciences de l'éducation*, la *Revue canadienne de psychoéducation*, les *Nouveaux Cahiers de la recherche en éducation* et la revue en ligne *Éducation et francophonie*. Ces revues scientifiques ont été retenues parce qu'elles ont une grande visibilité, couvrent un large éventail de thèmes du monde de l'éducation et constituent les meilleures sources pour cette étude. À cet égard, 250 articles ont été publiés au

total dans ces revues pendant la période visée par notre étude. Pour être sélectionnés, les travaux devaient forcément reposer sur une approche quantitative et ne pas utiliser uniquement des analyses relatives aux tests d'interdépendance (analyse factorielle ou en composante principale, analyse par grappes et analyse de cohérence interne) ainsi qu'aux tests non paramétriques. À cet égard, les articles retenus contiennent l'un ou l'autre des tests statistiques inférentiels suivants, pour répondre aux hypothèses ou aux objectifs de recherche: analyse de la variance univariée (tests *t* et ANOVA), à plan factoriel, multivariée (MANOVA) ou à mesures répétées (avec ou sans facteur inter-sujet), corrélation de Pearson, régression multiple ou logistique et test du khi-carré. Les travaux appuyés par des résultats qualitatifs ($n = 48$) n'ont évidemment pas été considérés. De même, les articles éditoriaux ($n = 16$) ou théoriques ($n = 63$), les recensions d'écrits ($n = 20$) ainsi que ceux où aucune analyse inférentielle n'était réalisée ($n = 20$) ont également été exclus. Quelques publications ont été écartées en raison d'un manque d'informations sur les paramètres essentiels au calcul de la puissance statistique ($n = 6$). Soulignons qu'en fonction des critères de sélection, ces 77 articles forment la *population* des travaux quantitatifs publiés par ces revues scientifiques durant la période couverte par cette étude et sont à cet égard forcément représentatifs des travaux de ce type publiés par ces revues.

Dans la mesure où l'on retrouve fréquemment plusieurs tests statistiques pour répondre aux objectifs ou aux hypothèses d'une recherche, la puissance statistique a été calculée pour chaque test portant directement sur les objectifs / hypothèses et ces résultats ont, par la suite, été compilés sous la forme d'une moyenne, de façon à ce que l'unité d'analyse soit une puissance moyenne pour chaque article de l'échantillon. Tout comme c'est le cas dans d'autres travaux du genre, les analyses statistiques secondaires ne font pas partie des analyses de puissance que nous présenterons (Cohen, 1962; Maddock et Rossi, 2001; Rossi, 1990; Sedlmeier et Gigerenzer, 1989).

3.2 Cueillette des données

Une fois sélectionné, chaque test inférentiel contenu à l'intérieur des articles portant sur la vérification des hypothèses ou des objectifs de recherche a été analysé en tenant compte des paramètres suivants: la taille de l'échantillon, le nombre de groupes (si pertinent), l'équivalence des groupes (oui ou non), le type de test statistique, le nombre de prédicteurs (dans le cas d'une régression) et le rejet ou l'acceptation de l'hypothèse nulle. De plus, nous avons noté, pour chaque article, l'absence ou la présence même minimale de propos décrivant une analyse de puissance statistique *a priori*, une justification du nombre de sujets, du seuil de signification adopté ou encore de la taille d'effet attendue. Enfin, nous avons noté la présence ou l'absence de toute préoccupation explicite du chercheur en regard des enjeux associés à la puissance statistique dans l'ensemble de l'article.

3.3 Calcul de la puissance statistique

Pour chaque test statistique, la puissance effective pour détecter un effet prédéterminé de petite, moyenne et grande taille à l'intérieur du devis a été calculée avec le logiciel *G*Power* (Faul et Erdfelder, 1992) pour *MS-DOS*. Nous avons donc calculé la puissance de chaque test à détecter trois effets hypothétiques de taille différente (petite, moyenne et grande) selon les balises proposées par Cohen (1988). Facile à utiliser et disponible gratuitement en ligne, ce logiciel permet de faire des analyses de puissance statistique *a priori* et *a posteriori* en insérant les valeurs des principaux déterminants de la puissance. En revanche, celui-ci ne prend pas en compte d'autres paramètres qui peuvent influencer l'indice de puissance comme la fidélité des mesures, le nombre de tests effectués, l'homogénéité de l'échantillon ou les variables confondantes. Nous aborderons quelques-uns de ces éléments et leur incidence sur la puissance dans la partie consacrée à la discussion. Le fonctionnement du logiciel est cohérent avec le cadre conceptuel de Cohen (1988). Un des avantages de ce logiciel est qu'il permet aussi d'estimer la puissance statistique pour des tests multivariés, comme l'analyse de variance à mesures répétées qui n'est pas couverte spécifiquement par les travaux de Cohen (1988, 1992). Dans la mesure où les techniques multivariées sont maintenant couramment employées à l'intérieur des travaux scientifiques en éducation, ce logiciel est un instrument indispensable pour l'analyse de la puissance statistique d'un devis de recherche. Enfin, dans le but de comparer, de manière optimale, nos résultats avec ceux des recensions antérieures, toutes les analyses ont été réalisées en fixant le seuil de signification à 0,05, en ne tenant pas compte du nombre de tests statistiques effectués par article et en considérant que l'hypothèse nulle était bidirectionnelle.

4. Résultats

4.1 Caractéristiques des devis

Avant de présenter les résultats d'analyse de puissance statistique, il est intéressant de documenter quelques paramètres des devis de recherche qui composent l'échantillon des études recensées. À partir des 77 études, un total de 172 tests statistiques portant directement sur les objectifs ou les hypothèses de recherche ont été analysés. Sur cette somme, 7,6 % était des tests *t*, 8,8 % des corrélations de Pearson, 33,9 % des analyses de variance simple ou de covariance, 20,5 % étaient des analyses multivariées de la variance, 18,1 % des régressions et 11,1 % des tests du khi-carré. En ce qui concerne la taille des échantillons, la distribution des valeurs est fortement décalée positivement, autant sur le plan de l'asymétrie que de l'aplatissement de la courbe (étendue de 18 à 9 448). Cela est dû au fait que quelques études ont été réalisées sur des échantillons de très grande taille. En effet, 13 % ($n = 10$) des devis portaient sur des échantillons de 1 000 observations et plus. La taille moyenne des échantillons est de 483 observations, alors que la médiane se situe à 139. Cela représente des valeurs très élevées, comparativement à ce qu'on retrouve dans les quelques travaux comparables et nous permet d'anticiper que ces études vont

certainement influencer à la hausse les indices moyens de puissance. Conséquentement, une analyse secondaire a été effectuée pour mieux apprécier l'impact de ces devis utilisant des échantillons de grande taille sur les moyennes de puissance obtenue pour notre étude. Cette analyse est présentée à la fin de la section des résultats.

Finalement, l'examen des articles montre que la presque totalité des chercheurs (96 %) ne sont pas préoccupés par les enjeux relatifs à la puissance statistique du devis de leur projet, du moins pas assez pour y faire référence dans le contenu écrit du texte publié. En effet, on ne retrouve des indications d'analyse de puissance *a priori* ou encore des allusions aux problèmes engendrés par le manque de puissance statistique qu'à l'intérieur de quelques travaux ($n = 3$). De même, soulignons qu'aucune étude ne fait mention d'éléments pouvant justifier la taille de l'échantillon sélectionné, le seuil alpha choisi ou encore la taille d'effet recherchée par les auteurs.

4.2 Analyse de puissance statistique

Dans le tableau 2 sont rapportées les distributions de la puissance statistique des 77 études selon les trois catégories de tailles d'effet préconisées par Cohen (1962, 1988). Ces résultats sont présentés exactement comme l'avait proposé Cohen (1962) et comme l'ont également fait la plupart des auteurs recensés dans le tableau 1. En outre, l'examen du tableau 2 permet de prendre connaissance des informations sur la forme des courbes de distribution de la puissance par les indices de moyenne et de médiane. Enfin, dans les deux dernières lignes de ce même tableau, on retrouve la proportion des études ayant une moyenne de puissance inférieure à 0,20, ce qui représente une puissance extrêmement faible, et la proportion d'études présentant une moyenne de puissance inférieure à 0,80, qui est le seuil de puissance optimal suggéré par Cohen (1988).

Un premier résultat concerne la capacité équivoque des devis étudiés à détecter des effets de petite taille. À cet égard, près de la moitié des articles recensés (48 %) ont une puissance statistique de moins de 0,20 pour détecter de tels effets et la puissance moyenne des devis est de 0,36. En revanche, on constate que seulement 16 % de ces études possèdent suffisamment de puissance pour égaliser ou dépasser le seuil de 0,80. Ces résultats indiquent qu'une proportion importante des devis ont de très faibles probabilités d'en arriver à des résultats significatifs dans la mesure où un effet de petite taille existe dans la population. En effet, dans 70 % des cas, cette probabilité de rejeter l'hypothèse nulle n'est même pas de une sur deux.

En ce qui concerne les effets de moyenne et grande taille, la situation est différente. Pour les effets de taille moyenne, la puissance moyenne des devis est de 0,78, ce qui est très près de la valeur optimale de puissance de 0,80 proposée par Cohen. En fait, la distribution des valeurs de puissance est concentrée vers les indices élevés, comme en témoigne la médiane située à 0,90.

Tableau 2

Distribution des fréquences et fréquences cumulatives de la puissance statistique moyenne de 77 devis de recherche pour détecter des effets de petite, moyenne et grande taille (hypothèses bidirectionnelles et $\alpha = 0,05$)

Puissance statistique	Effet de petite taille		Effet de moyenne taille		Effet de grande taille	
	Fréquence	Cumul (%)	Fréquence	Cumul (%)	Fréquence	Cumul (%)
0,99 – 1,00	4	100 %	25	100 %	46	100 %
0,95 - 0,98	2	95 %	8	68 %	9	40 %
0,90 – 0,94	1	92 %	6	57 %	7	29 %
0,80 – 0,89	5	91 %	4	49 %	6	19 %
0,70 – 0,79	4	84 %	8	44 %	5	12 %
0,60 – 0,69	3	79 %	5	34 %	1	5 %
0,50 – 0,59	4	75 %	11	27 %	1	4 %
0,40 – 0,49	1	70 %	1	13 %	1	3 %
0,30 – 0,39	5	69 %	6	12 %		
0,20 – 0,29	11	62 %	2	4 %	1	1 %
0,10 – 0,19	28	48 %	1	1 %		
0,05 – 0,09	9	12 %				
Indices de tendance centrale	Effet de petite taille		Effet de taille moyenne		Effet de grande taille	
Moyenne	0,36		0,78		0,93	
Médiane	0,20		0,90		1,00	
Proportion d'études dont la puissance est :						
< 0,20	48 %		1 %		0 %	
< 0,80	84 %		44 %		12 %	

Les devis ont donc, pour la plupart, une puissance suffisante pour détecter des effets de taille moyenne. Cependant, il faut souligner que 44 % d'entre eux présentent toujours une puissance statistique en deçà du seuil recommandé de 0,80, et que 27 % ont une puissance moyenne inférieure à 0,50. Enfin, la puissance des études pour détecter des effets de grande taille est largement au-dessus de 0,80, avec une moyenne de 0,93. Avec seulement 12 % des devis sous la barre de 0,80, il semble que la majorité des recherches possèdent des devis amplement robustes pour rejeter l'hypothèse nulle lorsqu'un effet de grande taille est présent dans la population.

4.3 Impact de la taille des échantillons

Nous avons constaté précédemment qu'une des caractéristiques de l'échantillon d'articles recensés est la présence d'une minorité de devis (13 %) constitués d'échantillons de très grande taille. Lorsque les devis sont distribués selon la taille de l'échantillon, ces échantillons exceptionnels se trouvent très loin de la moyenne des études et représentent des valeurs extrêmes de la distribution. À cet égard, nous

avons repris les analyses de fréquence en excluant cette fois-ci les devis qui comptent plus de 500 observations. Les moyennes de puissance statistique obtenues avec cette distribution plus homogène sont de 0,22 pour les effets de petite taille ; 0,72 pour les effets de taille moyenne et 0,92 pour les effets de grande taille. La forme des distributions pour chaque taille d'effet reste donc sensiblement la même avec ce sous-échantillon d'études, avec une asymétrie positive prononcée pour la distribution relative à la petite taille d'effet et une asymétrie négative pour la distribution relative à l'effet de grande taille.

5. Discussion

L'objectif principal de cette étude était de faire le point sur la puissance statistique des devis de recherche du Canada francophone en éducation. Mis à part quelques cas d'exception, l'analyse des données recueillies lors de l'examen des textes scientifiques montre d'entrée de jeu que les chercheurs ne semblent pas se préoccuper des enjeux liés à la puissance statistique. À tout le moins, cette préoccupation n'apparaît pas à l'écrit. En effet, à la lecture des articles, nous n'avons trouvé aucune trace de justification ou d'explication relatives au choix du seuil de signification, à la direction des tests d'hypothèse, au nombre de sujets sélectionnés et, encore moins, à propos de la taille d'effet recherchée. Ces premiers constats sont troublants, d'une part parce qu'ils mettent en lumière l'absence généralisée de propos scientifiques sur les éléments associés à la puissance statistique dans la présentation des devis de recherche publiés dans le domaine de l'éducation et, d'autre part, parce que ces constats correspondent aux conclusions tirées par les analyses de ce type menées depuis 40 ans (Brewer, 1972 ; Cohen, 1962 ; Maddock et Rossi, 2001 ; Rossi, 1990). À cet égard, il est étonnant de constater qu'en dépit d'une quantité substantielle et récurrente d'écrits sur le sujet, la grande majorité des auteurs de notre champ de recherche ne portent aucune attention apparente à ces constituants fondamentaux des devis de recherche quantitatifs. Vue sous cet angle, la situation semble n'avoir guère évolué.

En contrepartie, notre analyse révèle que la puissance moyenne des devis de recherche étudiés est supérieure à celle que l'on trouve dans les recherches de même type menées depuis les travaux pionniers de Cohen. En effet, comparées à la seule étude disponible sur la puissance statistique dans le domaine de l'éducation conduite par Brewer en 1972, les valeurs moyennes de puissance ont augmenté d'environ 20 points, et ce, pour les trois catégories de taille d'effet. Cette tendance se maintient aussi lorsque l'on compare les résultats de notre étude avec ceux des autres recherches sur la puissance présentées précédemment. Cela dit, il appert que cet écart important entre nos résultats et ceux des analyses de puissance provenant des domaines autres que celui de l'éducation provient essentiellement de différences notables sur le nombre moyen d'observations des recherches sélectionnées. En effet, pour les travaux qui précisent le nombre moyen de sujets (3 sur 8), l'écart est pour le moins frappant. Alors que le nombre moyen d'observations est

de 483 dans notre étude, Cohen (1962) en rapporte 68, Magid et ses collaborateurs (1987) indiquent un nombre médian de 41 et Haase (1974) note un nombre moyen de 28 par étude recensée.

Il est donc clair que la présence d'échantillons de très grande taille parmi les études recensées entre 2004 et 2006 explique, en grande partie, l'amélioration des valeurs moyennes de puissance statistique. Cependant, les devis qui disposent de tels échantillons ne représentent qu'une minorité des travaux sélectionnés pour notre étude et, lorsque nous effectuons l'exercice de les exclure des calculs de puissance, nous retrouvons des valeurs de puissance relativement proches des résultats classiques sur le sujet. Il faut donc nuancer cette apparente amélioration de la puissance moyenne des devis en éducation en tenant compte du fait qu'une minorité de chercheurs ont eu accès à un bassin important de sujets et que cette opportunité, d'une part, n'est certainement pas représentative de la réalité de tous les chercheurs et, d'autre part, n'est pas nécessairement justifiée.

Vue selon la perspective de la taille de l'échantillon, l'absence d'analyse de la puissance statistique nécessaire lors de la planification des devis de recherche conduit forcément le chercheur vers deux avenues. La première avenue consiste à accumuler un nombre trop important de sujets en regard de l'objectif de l'étude. En effet, on peut se demander si un échantillon de 10 000 sujets est nécessaire lorsqu'on désire détecter un effet de taille moyenne en utilisant une technique d'analyse statistique de la variance simple. Cette démesure méthodologique est tout aussi questionnable que la présence d'un nombre insuffisant de sujets dans un devis de recherche. À titre d'exemple, une des 77 études sélectionnées effectuée des tests de corrélation ainsi que des tests du khi-carré à partir d'un échantillon de 1 737 sujets. En se basant sur ces informations et les tableaux de puissance de Cohen (1992), on remarque que le nombre optimal de sujets nécessaires pour détecter un effet de petite taille avec un seuil alpha de 0,05 est de $n = 783$ pour les analyses de corrélation et de $n = 1 090$ pour les analyses de khi-carré avec trois degrés de liberté. Pour cette étude, il y a donc un effort supplémentaire, pour le moins discutable, de cueillette de données auprès de 653 sujets. En outre, si ces chercheurs avaient déterminé au préalable qu'un effet de taille moyenne était recherché, les tailles d'échantillon prescrites auraient chuté à $n = 85$ pour les corrélations et à $n = 121$ pour le même khi-carré. En considérant que l'ensemble de ces données n'a été collecté que pour cette étude bien précise, que des coûts importants sont reliés à la sollicitation des milieux, à la prise de données sur le terrain ainsi qu'au traitement des données, on peut croire qu'une économie substantielle aurait pu être ici réalisée si les chercheurs avaient calculé *a priori* la puissance statistique de ce devis.

La seconde voie possible est de recruter un nombre trop faible de sujets et, donc, manquer de puissance statistique. Cette situation embarrassante est la plus courante. Comme nous l'avons vu précédemment, dans le cas d'une puissance de 0,50, les probabilités de détecter un effet en termes de résultats significatifs sont les

mêmes que de tirer une pièce de monnaie en l'air. Bien que le portrait général de la puissance statistique semble s'être sensiblement amélioré, une importante proportion d'études en éducation montrent encore une puissance statistique insuffisante pour détecter des effets de petite taille (84 %) et même des effets de taille moyenne (44 %). Ces données devraient à elles seules préoccuper les chercheurs, car il semble que la majorité des phénomènes étudiés en éducation se situent entre ces deux catégories (Cohen, 1988 ; Haase, 1974 ; Rossi, 1990). À ce sujet, Rossi (1990) affirme qu'en plus d'exposer le chercheur aux erreurs de Type II, une puissance statistique insuffisante dans un domaine de recherche se traduit inévitablement par une prolifération d'erreurs de Type I. Imaginons un domaine de recherche dans lequel la taille d'effet du phénomène observé est nulle. La probabilité d'arriver à des résultats significatifs est forcément égale au seuil *alpha*. Et dans la mesure où les journaux publient presque exclusivement des recherches portant des résultats significatifs, on ne peut réfuter la possibilité que bon nombre de rejets de l'hypothèse nulle en contexte d'un manque flagrant de puissance statistique sont en fait des erreurs de Type I. Ainsi, le manque de puissance statistique peut également expliquer les résultats de recherche contradictoires dans certains domaines d'étude.

Peu importe qu'une recherche possède beaucoup trop ou trop peu de puissance statistique, les éléments de discussions que nous avançons sont autant d'invitations lancées aux chercheurs pour planifier adéquatement leur devis de recherche en considérant les implications reliées à l'équilibre à atteindre entre l'importance accordée au seuil *alpha* et celle qu'on devrait également accorder au seuil *beta*. Pour ce faire, il est primordial que les chercheurs en éducation se penchent en premier lieu sur la taille de l'effet qu'ils désirent détecter. Cette réflexion est certes ardue et encore peu balisée par les écrits scientifiques, mais elle reste essentielle dans le processus de conception d'un devis adéquat. Personne ne trouverait judicieux de tuer une mouche avec un canon ou encore d'assommer un éléphant avec un cure-dent. Pourtant, ces analogies représentent assez bien les armatures extrêmes dont sont pourvus plusieurs devis afin de répondre aux objectifs ou aux hypothèses de recherche.

Un problème courant auquel sont confrontés une majorité de chercheurs est la difficulté de recruter suffisamment de sujets pour assurer une puissance statistique adéquate. Souvent tributaire de contraintes économiques ou contextuelles, le chercheur doit tout de même composer avec cette réalité. À cet égard, il est important de savoir qu'il existe d'autres solutions pour maximiser la puissance statistique d'un devis que d'augmenter la taille de l'échantillon ou de diminuer le seuil *alpha* lors des analyses. En fait, toutes les stratégies visant à réduire la variabilité des observations concourent à l'augmentation de la puissance statistique. Sur le plan des participants, ces stratégies visent à réduire l'hétérogénéité de l'échantillon. On peut penser à une application rigoureuse des critères de sélection des sujets ainsi qu'à l'utilisation de covariables dans certaines analyses dans le but

de contrôler les principales sources de variation entre les participants. De la même manière, sur le plan de la mesure, il est reconnu que les sources d'erreur métrique affectent également la puissance d'un devis. Bien que la relation entre la puissance statistique et la fidélité ne soit pas une fonction mathématique directe, plusieurs auteurs concluent qu'il est primordial d'utiliser les instruments les plus précis possibles pour assurer une puissance optimale (Williams et Zimmerman, 1989; Williams, Zimmerman et Zumbo, 1995). Ainsi, il est important de choisir des instruments de mesure possédant de bonnes qualités métrologiques, car plus le coefficient de fidélité d'un instrument est élevé, plus la portion d'erreur de la variance du score est réduite, et cela a un effet à la hausse sur la puissance statistique du devis. En effet, les fluctuations de la mesure causées par le manque de fidélité d'un instrument augmentent l'étendue des intervalles de confiance lors de l'estimation des paramètres et réduisent ainsi les probabilités de détecter des effets significatifs. Il faut mentionner aussi que d'autres éléments relatifs à la mesure, comme le nombre d'items et les indices de validité, peuvent également avoir un impact sur la puissance statistique d'un devis. L'auteur renvoie le lecteur à l'article de Williams et Zimmerman (1989) pour un tour d'horizon sur la question. Voilà donc quelques éléments sur lesquels les chercheurs ont un certain pouvoir et qui jouent un rôle prépondérant dans la capacité d'un devis à détecter adéquatement l'effet envisagé. De plus, certaines modalités concernant les analyses statistiques favorisent une puissance statistique optimale. Chaque fois que le contexte s'y prête, Rossi (1990) recommande d'avoir systématiquement recours aux techniques d'analyse multivariées, et ce, dans le but de limiter le nombre de tests effectués sur les mêmes variables. En effet, Sedlmeier et Gigerenzer (1989) ont montré que les ajustements du seuil *alpha*, habituellement effectués lors de la multiplication de tests réalisés sur les mêmes variables, diminuent fortement la puissance statistique de ces analyses. En outre, le principe de parcimonie dans le nombre de variables à considérer à l'intérieur d'un modèle multivarié (Hair, Anderson, Tatham et Black, 1998) ainsi que la diminution du degré de liberté dans le test du khi-carré par la réduction du nombre de catégories sont autant de stratégies d'ordre technique facilement applicables qui visent à maximiser la puissance statistique d'un devis.

Notre étude présente quelques limites inhérentes aux choix méthodologiques. En effet, nous avons choisi de couvrir deux années récentes de publication, ce qui ne nous permet pas de conclure que la tendance à la hausse des valeurs moyennes de puissance que nous observons est un phénomène ponctuel ou bien une caractéristique des recherches contemporaines en éducation. À ce titre, bien que la recherche de Brewer (1972) ait couvert une période d'une dizaine d'années, ses résultats étaient similaires à ceux obtenus par les autres chercheurs ayant couvert des périodes de publication plus courtes. À tout le moins, on peut penser qu'un intervalle plus étendu inclurait davantage d'études et permettrait de préciser la tendance. Rappelons toutefois que le but de la présente étude était de faire état de la situation actuelle des devis de recherche en éducation, et non de décrire de

manière exhaustive l'évolution de la puissance statistique de ces devis depuis les travaux de Brewer (1972). Enfin, tout comme dans les autres travaux de la même catégorie, les valeurs moyennes de puissance sont certainement surestimées dans la mesure où les méthodes actuelles de calcul de la puissance ne prennent pas en compte les sources internes ou externes de variabilité. Comme cela est rarement le cas, on doit donc considérer ces valeurs comme étant optimales, et que la puissance moyenne réelle est probablement plus faible.

6. Conclusion

Les chercheurs en éducation du Canada francophone ont encore beaucoup à faire pour assurer une puissance statistique adéquate aux devis des travaux scientifiques qui sont publiés. Les résultats ont montré que très peu de progrès ont été réalisés en 40 ans en cette matière. Ce constat est troublant et inquiétant, car il suggère qu'une proportion importante des résultats de recherche dans notre domaine pourrait être remise en question en raison d'une trop faible puissance statistique des devis employés pour répondre aux objectifs et aux hypothèses de recherche. Il est clair que des travaux comme celui-ci sont nécessaires pour poursuivre la sensibilisation de la communauté scientifique en matière de puissance statistique. En effet, les enjeux liés à la puissance statistique dépassent les considérations méthodologiques : ils sont directement associés à la crédibilité et à la validité des résultats de recherche en éducation. À cet égard, il pourrait être souhaitable que les éditeurs de revue scientifique demandent aux auteurs de donner des informations minimales sur la puissance statistique de leurs travaux.

En terminant, il serait intéressant de reprendre la démarche de cet article à l'intérieur des champs spécifiques de la recherche en éducation (pédagogie, psychoéducation, orientation professionnelle, andragogie). En effet, cette étude ne permet pas d'établir une distinction entre les différentes sphères du monde de la recherche en éducation. En ce sens, des études portant sur des domaines spécifiques d'intervention éducative pourraient certainement aider à préciser davantage le portrait de la puissance statistique des devis de recherche en éducation.

ENGLISH TITLE • The statistical power of research designs in education

SUMMARY • This article examines the mean statistical power of recent articles published in French Canadian scientific journals dedicated to research in education. 77 studies published in 4 different journals between 2004 and 2006 were analyzed with G*Power to establish their potential to detect small, medium, and large effects. Overall, results show that little progress has been made since the founding work of Cohen (1962) regarding statistical power in the field of educational research in the francophone scientific community. The author gives concrete actions for researchers to improve the statistical power of future work.

KEY WORDS • statistical power, quantitative research, methodology, inferential statistics, internal validity.

TÍTULO EN ESPAÑOL • Estudio sobre la potencia estadística de los diseños de investigación en educación

RESUMEN • El presente artículo tiene por objetivo analizar la potencia estadística promedio de los diseños de investigación en educación, publicados en los últimos años en revistas francocanadienses. Fueron analizados 77 artículos publicados entre 2004 y 2006 en cuatro revistas científicas francocanadienses con el software *G*Power* para determinar su capacidad para detectar efectos de tamaño pequeño, medio o grande. Los resultados muestran que la investigación francocanadiense en educación realizó pocos adelantos reales sobre el tema de la potencia estadística desde los trabajos de Cohen (1962). Se recomiendan algunas pistas concretas para mejorar la potencia estadística de los diseños de investigación.

PALABRAS CLAVES • potencia estadística, investigación cuantitativa, diseño de investigación, estadísticas inferenciales, validez interna.

Références

- *Armand, F. (2005). Capacités métalinguistiques d'élèves immigrants nouvellement arrivés en situation de grand retard scolaire. *Revue des sciences de l'éducation*, 31(2), 441-469.
- *Beaumont, C., Royer, É., Bowen, F. et Bertrand, R. (2004). L'adaptation psychosociale des élèves en trouble de comportement agissant comme médiateurs. *Revue des sciences de l'éducation*, 30(3), 555-578.
- *Beer-Toker, M. et Gaudreau, A. (2006). Représentations, attitudes et pratiques de littératie chez des élèves allophones : construction d'un outil de dépistage des difficultés en matière de littératie. *Revue des sciences de l'éducation*, 32(2), 345-376.
- *Blanchard, C., Pelletier, L., Otis, N. et Sharp, E. (2004). Rôle de l'autodétermination et des aptitudes scolaires dans la prédiction des absences scolaires et l'intention de décrocher. *Revue des sciences de l'éducation*, 30(1), 105-123.
- *Bouchamma, Y., Daniel, I. et Moisset, J.-J. (2004). Les causes et la prévention de la violence en milieu scolaire haïtien : ce qu'en pensent les directions d'école. *Éducation et francophonie*, 32(1), 87-101.
- *Boudreau, M., Saint-Laurent, M. et Giasson, J. (2006). La littératie familiale et les habiletés en conscience phonologique des enfants de maternelle. *Éducation et francophonie*, 34(2), 189-212.
- Bourque, J., Blais, J. G. et Larose, F. (2009). L'interprétation des tests d'hypothèses : p , la taille de l'effet et la puissance. *Revue des sciences de l'éducation*, 35(1), 211-225.
- *Bowen, F., Chouinard, R. et Janosz, M. (2004). Modèle des déterminants des buts de maîtrise chez les élèves. *Revue des sciences de l'éducation*, 30(1), 49-70.
- Brewer, J. K. (1972). On the power of statistical tests in the American Educational Research. *Journal of american educational research journal*, 9(3), 391-401.
- *Brodeur, M. (2006). Prévention des difficultés d'apprentissage en lecture : l'effet différencié d'un programme implanté par des enseignantes de maternelle chez leurs élèves. *Éducation et francophonie*, 34(2), 56-83.
- *Brodeur, M., Gosselin, C., Legault, F., Deaudelin, C., Mercier, J. et Vanier, N. (2005). Prévention des difficultés d'apprentissage en lecture chez les enseignants de maternelle. *Revue des sciences de l'éducation*, 31(1), 33-54.

- *Cantin, S. et Boivin, M. (2005). Les parents et le groupe de pairs comme sources potentielles de soutien social lors de la transition scolaire primaire-secondaire. *Revue de psychoéducation*, 34(1), 1-19.
- *Cavanagh, M. (2006). Validation d'un programme d'intervention : pour la cohérence des écrits argumentatifs au primaire. *Revue des sciences de l'éducation*, 32(1), 159-182.
- *Charlebois, P., Vitaro, R., Normandeau, S., Lévesque, J. et Rondeau, N. (2006). Illustration des principes et des procédures d'implantation et d'évaluation d'un programme de prévention de type ciblé. *Revue de psychoéducation*, 35(1), 65-94.
- *Chouinard, R., Plouffe, C. et Roy, N. (2004). Caractéristiques motivationnelles des garçons du secondaire en difficulté d'apprentissage ou en trouble de la conduite. *Revue des sciences de l'éducation*, 30(1), 143-162.
- *Clarisse, R., LeFloc'h, N., Testu, F. et Fournier, C. (2006). Variations de l'attention et du sommeil d'enfants de 7-8 ans en fonction des modes de transport scolaire utilisés. *Revue de psychoéducation*, 35(1), 141-153.
- Cohen, J. (1962). The statistical power of abnormal social psychological research: a review. *Journal of abnormal and social psychology*, 65(3), 145-153.
- Cohen, J. (1988). *Statistical power for the behavioral sciences* (2^e édition). New York, New York: Academic Press.
- Cohen, J. (1990). Things I have learned (so far). *American psychologist*, 45(12), 1304-1312.
- Cohen, J. (1992). A power primer. *Psychological bulletin*, 112(1), 155-159.
- Cohen, J. (1994). The earth is round ($p < 0,05$). *American psychologist*, 49(12), 997-1003.
- *Cormier, P. (2006). Connaissance du nom des lettres chez des enfants francophones de 4, 5 et 6 ans. *Éducation et francophonie*, 34(2), 5-27.
- *Coslin, P. G. (2004). Relation aux parents et violences scolaires. *Éducation et francophonie*, 32(1), 126-137.
- *Cossette, M.-C., Potvin, P., Marcotte, D., Fortin, L., Royer, É. et Leclerc, D. (2004). Le risque de décrochage scolaire et la perception du climat de classe chez les élèves du secondaire. *Revue de psychoéducation*, 33(1), 117-136.
- *Denis, É., Malcuit, G. et Pomerleau, A. (2005). Évaluation des impacts de l'initiative communautaire 1, 2, 3 Go! sur le développement et le bien-être des tout-petits et de leur famille. *Éducation et francophonie*, 33(2), 44-66.
- *Denoncourt, I., Bouffard, T., Dubois, V. et McIntyre, M. (2004). Relations entre les facteurs du profil motivationnel d'élèves de sixième année du primaire et leurs anticipations envers le secondaire. *Revue des sciences de l'éducation*, 30(1), 71-89.
- *Déry, M., Lapalme, M. et Yergeau, É. (2005). Les services scolaires et sociaux offerts aux élèves du primaire présentant des troubles de comportement. *Nouveaux cahiers de la recherche en éducation*, 8(2), 37-46.
- *Desbiens, J.-F. (2006). Enseigner et apprendre en classe multiprogramme : un débat souterrain, des enjeux d'actualité. *Nouveaux cahiers de la recherche en éducation*, 9(1), 91-110.
- *Deslandes, R. et Bertrand, R. (2004). Motivation des parents à participer au suivi scolaire de leur enfant au primaire. *Revue des sciences de l'éducation*, 30(2), 411-433.

- *Deslandes, R. et Jacques, M. (2004). Relations famille-école et l'ajustement du comportement socioscolaire de l'enfant à l'éducation préscolaire. *Éducation et francophonie*, 32(1), 172-200.
- *Dionne, A.-M., Saint-Laurent, L. et Giasson, J. (2005). Analyse des interactions entre le parent faiblement scolarisé et son enfant lors de la lecture. *Revue de psychoéducation*, 34(2), 175-190.
- *Dumont, M., Pronovost, J. et Leclerc, D. (2004). Les stratégies adaptatives des adolescents : comparaison d'un groupe scolaire et d'un groupe desservi en Centres jeunesse. *Revue de psychoéducation*, 33(1), 137-155.
- *Dupeyrat, C. et Mariné, C. (2004). Conceptions de l'intelligence, orientations de buts et stratégies d'apprentissage chez les adultes en reprise d'études. *Revue des sciences de l'éducation*, 30(1), 27-48.
- Faul, F. et Erdfelder, E. (1992). *G*Power: a priori, post hoc and compromise power analysis for MS-DOS*. Programme informatique. Bonn, Allemagne : Université de Bonn, département de psychologie.
- Fisher, R. A. (1956). *Statistical methods and scientific inference*. Edinburgh, Scotland : Oliver et Boyd.
- *Fortin, L., Marcotte, D., Royer, É. et Potvin, P. (2005). Facteurs personnels, scolaires et familiaux différenciant les garçons en problèmes de comportement du secondaire qui ont décroché ou non de l'école. *Nouveaux cahiers de la recherche en éducation*, 8(2), 79-88.
- *Foulin, J.-N. et Pacton, S. (2006). La connaissance du nom des lettres : précurseur de l'apprentissage du son des lettres. *Éducation et francophonie*, 34(2), 28-55.
- *Gagné, M.-H., Lavoie, F. et Hébert, M. (2004). Exploration des rapports entre différentes expériences avec la violence dans la vie des jeunes québécois(es). *Revue de psychoéducation*, 33(1), 41-60.
- *Galand, B. (2004). Le rôle du contexte scolaire et de la démotivation dans l'absentéisme des élèves. *Revue des sciences de l'éducation*, 30(1), 125-142.
- *Galand, B., Philippot, P., Petit, S., Born, M. et Buidin, G. (2004). Regards croisés sur les phénomènes de violence en milieu scolaire : élèves et équipes éducatives. *Revue des sciences de l'éducation*, 30(3), 465-486.
- *Gamba, C., Martinet, C. et Saada-Robert, M. (2006). Microgénèses comparées en lecture interactive à 4 ans : co-construction de la langue écrite entre élèves et enseignants. *Éducation et francophonie*, 34(2), 213-234.
- *Gendron, M., Royer, É., Bertrand, R. et Potvin, P. (2005). Les troubles du comportement, la compétence sociale et la pratique d'activités physiques chez les adolescents. *Revue des sciences de l'éducation*, 31(2), 211-233.
- *Gendron, M., Royer, É., Bertrand, R., Potvin, P. et Frenette, E. (2006). Relation entre l'adaptation psychosociale, les habitudes de vie et les habiletés sociales chez les adolescents présentant des troubles du comportement. *Revue de psychoéducation*, 35(2), 229-205.
- Haase, R. F. (1974). Power analysis of research in counsellor education. *Counselor education and supervision*, 14(2), 124-132.

- Hair, J. F., Anderson, R. E., Tatham, R. L. et Black, W. C. (1998). *Multivariate data analysis* (5^e édition). Englewood Cliffs, New Jersey: Prentice Hall.
- *Hébert, M. (2004). Les cercles littéraires entre pairs en première secondaire: étude des relations entre les modalités de lecture et de collaboration. *Revue des sciences de l'éducation*, 30(3), 605-630.
- *Huet, N. et Escribe, C. (2004). Croyances épistémiques, buts d'accomplissement de soi et engagement dans l'utilisation d'un média électronique chez des étudiants. *Revue des sciences de l'éducation*, 30(1), 177-196.
- *Jalbert, J. et Pagani, L. (2005). La composante parentale d'un programme d'éveil aux mathématiques destiné aux enfants de milieu socioéconomiquement faible au cours de la prématernelle représente-t-elle une valeur ajoutée? *Revue de psychoéducation*, 34(2), 301-329.
- *Jutras, F., Joly, J., Legault, G. A. et Desaulniers, M.-P. (2005). L'intervention professionnelle en enseignement: les conceptions de la profession chez le personnel du primaire et du secondaire. *Revue des sciences de l'éducation*, 31(3), 563-583.
- Kazdin, A. E. et Bass, D. (1989). Power to detect differences between alternative treatments in comparative psychotherapy outcome research. *Journal of consulting and clinical psychology*, 57(1), 138-147.
- *Lafortune, D., Laurier, C. et Gagnon, F. (2004). Prévalence et facteurs associés à la prescription de médicaments psychotropes chez les sujets placés en Centre jeunesse. *Revue de psychoéducation*, 33(1), 157-176.
- *Lancôt, N. (2005). Liens entre l'inadaptation scolaire des adolescents et des adolescentes judiciarisés et leur adaptation sociale et personnelle à l'âge adulte. *Nouveaux cahiers de la recherche en éducation*, 8(2), 89-100.
- *Landry, R., Deveau, K. et Allard, R. (2006). Vitalité ethnolinguistique et construction identitaire: le cas de l'identité bilingue. *Éducation et francophonie*, 34(1), 54-81.
- *Le Corff, Y., Toupin, J. et Pauzé, R. (2005). Les troubles de comportement à l'adolescence: description, adaptation et évolution. *Nouveaux cahiers de la recherche en éducation*, 8(2), 69-78.
- *Lemétayer, F. et Kraemer, D. (2005). Contexte scolaire et estime de soi à l'adolescence: étude comparative entre adolescents non déficients et adolescents présentant une déficience intellectuelle légère. *Revue de psychoéducation*, 34(1), 57-70.
- *Lentillon, V. et Trottin, B. (2005). Relations éducatives en éducation physique et sportive et perceptions chez des collégiennes et des collégiens. *Éducation et francophonie*, 32(3), 57-72.
- Maddock, J. E. et Rossi, J. S. (2001). Statistical power of articles published in three health psychology related journals. *Health psychology*, 20(1), 76-78.
- Magid, A., Mazen, M., Hemmasi, M. et Lewis, M. F. (1987). Assessment of statistical power in contemporary strategy research. *Strategic management journal*, 8, 403-410.
- *Maïano, C., Ninot, G., Bilard, J., Benattar, B. et Bégarie, J. (2004). Effet à moyen terme de rencontres alternées en basket-ball sur l'auto-illusion de compétence physique d'adolescents présentant une déficience intellectuelle moyenne et légère. *Revue de psychoéducation*, 33(2), 305-320.

- *Maltais, C. (2005). Relation entre les types de services de garde et le développement du langage chez les enfants du préscolaire. *Éducation et francophonie*, 33(2), 207-223.
- *Marcotte, D., Cournoyer, M., Gagné, M.-È. et Bélanger, M. (2005). Comparaison des facteurs personnels, scolaires et familiaux associés aux troubles intériorisés à la fin du primaire et au début du secondaire. *Nouveaux cahiers de la recherche en éducation*, 8(2), 57-67.
- *Marcotte, J., Fortin, L., Cloutier, R., Royer, É. et Marcotte, D. (2005). Évolution de l'engagement parental auprès des élèves en difficulté de comportement et des élèves ordinaires au début du secondaire. *Nouveaux cahiers de la recherche en éducation*, 8(2), 47-56.
- Marks, H. M. (2003). Rigorous uncertainty: why RA Fisher is important. *International journal of epidemiology*, 32(6), 932-937.
- *Martin, F., Morcillo, A. et Blin, J.-F. (2004). Le vécu émotionnel des enseignants confrontés à des perturbations scolaires. *Revue des sciences de l'éducation*, 30(3), 579-604.
- *Martinet, C. et Rieben, L. (2006). Copie de mots, connaissance des lettres et conscience phonémique: une étude longitudinale chez des enfants de 5 ans. *Éducation et francophonie*, 34(2), 103-124.
- *Masny, D. (2006). Le développement de l'écrit en milieu de langue minoritaire: l'apport de la communication orale et des habiletés métalinguistiques. *Éducation et francophonie*, 34(2), 125-148.
- *Mayer, M., Dufour, S., Lavergne, C., Girard, M. et Trochmé, N. (2006). Structures familiales, paternité et négligence: des réalités à revisiter. *Revue de psychoéducation*, 35(1), 155-176.
- *Morcillo, A., Martin, F. et Jeunier, B. (2004). Approche psycho-éducative de la déviance scolaire. *Éducation et francophonie*, 32(1), 327-340.
- Neyman, J. (1942). Basic ideas and some recent results of the theory of testing statistical hypothesis. *Journal of the Royal statistical society*, 105, 292-327.
- Orme, J. G. et Tolman, R. M. (1986). The statistical power of a decade of social work education research. *Social service review*, 60, 619-632.
- *Pakzad, S. et Rogé, B. (2005). Estime de soi et dyslexie développementale. *Revue de psychoéducation*, 34(1), 71-92.
- *Paquette, G., Pauzé, R. et Joly, J. (2006). Caractéristiques sociofamiliales et personnelles qui permettent de distinguer les filles et les garçons présentant un trouble des conduites. *Revue de psychoéducation*, 35(2), 251-275.
- *Parent, S., Montésinos-Gelet, I., Séguin, J. R., Zelazo, P. D. et Tremblay, R. E. (2006). La contribution de la diversité des expériences littéraires préscolaires aux habiletés émergentes en littéracie. *Éducation et francophonie*, 34(2), 168-188.
- *Pasa, L., Creuzet, V. et Fijalkow, J. (2006). Écriture inventée: pluralité des traitements et variabilité selon la structure syllabique. *Éducation et francophonie*, 34(2), 84-102.
- *Quiroga, C., Janosz, M. et Marcotte, D. (2006). Les sentiments dépressifs à l'adolescence: un facteur de risque différentiel du décrochage scolaire chez les filles et les garçons de milieu défavorisé. *Revue de psychoéducation*, 35(2), 277-300.
- *Rasseneur, D. et Born, M. (2004). Facteurs familiaux de la délinquance juvénile féminine. *Revue de psychoéducation*, 33(2), 321-333.

- *Rey, V., Sabater, C. et Prost, V. (2006). Influence de l'apprentissage de l'écrit lors d'une tâche de répétition de mots et de logatomes, avec lecture labiale et sans lecture labiale, chez des enfants de 4 à 8 ans. *Éducation et francophonie*, 34(2), 149-167.
- Rossi, J. S. (1990). Statistical power of psychological research: what have we gained in 20 years? *Journal of consulting and clinical psychology*, 58(5), 646-656.
- *Saint-Jacques, M.-C., Cloutier, R., Pauzé, R., Simard, M. et Gagné, M.-H. (2004). La spécificité de la problématique des jeunes suivis en Centre jeunesse provenant de familles recomposées. *Revue de psychoéducation*, 33(2), 335-358.
- Sedlmeier, P. et Gigerenzer, G. (1989). Do studies of statistical power have an effect on power of studies? *Psychological bulletin*, 105(2), 309-316.
- *Sénécal, M., Hébert, M., Lavoie, F., Tremblay, R. E. et Vitaro, F. (2006). Fréquence et corrélatés des expériences de harcèlement sexuel chez des jeunes adultes francophones du Québec. *Revue de psychoéducation*, 35(2), 355-375.
- *Sylla, N. et DeVos, L. (2006). Facteurs déterminant l'application de nouvelles tâches d'enseignement dans le secondaire supérieur. *Revue des sciences de l'éducation*, 32(2), 377-394.
- *Toczek, M.-C., Michinow, E. et Michinow, N. (2006). Coopérer à distance... un contexte équitable pour les filles et les garçons? *Revue des sciences de l'éducation*, 32(2), 261-282.
- *Toupin, J., Déry, M. et Pauzé, R. (2005). Les services des Centres jeunesse offerts aux adolescents ayant des troubles de comportement II: évolution des jeunes et des familles une année plus tard. *Revue de psychoéducation*, 34(2), 215-231.
- *Toupin, J., Pauzé, R. et Déry, M. (2005). Les services des Centres jeunesse offerts aux adolescents ayant des troubles de comportement I: associations avec les difficultés des jeunes et des familles. *Revue de psychoéducation*, 34(2), 191-213.
- *Tourigny, M., Lavoie, F., Vézina, J. et Pelletier, V. (2006). La violence subie par des adolescentes dans leurs fréquentations amoureuses: incidence et facteurs associés. *Revue de psychoéducation*, 35(2), 323-354.
- *Tremblay, D.-G. (2005). Les communautés de pratique: une analyse différenciée selon le sexe de ce mode d'apprentissage. *Éducation et francophonie*, 32(3), 140-164.
- *Trudel, M., Blais, C. et Marceau, R. (2005). Évolution des processus d'intégration sociale des élèves en fonction du profil de santé mentale au début du primaire. *Nouveaux cahiers de la recherche en éducation*, 8(2), 7-16.
- *Turcotte, D. et Lamonde, G. (2004). La violence à l'école primaire: les auteurs et les victimes. *Éducation et francophonie*, 32(1), 15-37.
- *Venet, M., Bureau, J.-F., Dubois-Comptois, K., Gosselin, C. et Capuano, F. (2006). L'évaluation représentationnelle des modèles opérationnels internes de la relation d'attachement à l'âge préscolaire. *Revue de psychoéducation*, 35(1), 121-140.
- *Verlaan, P., Charbonneau, M.-N. et Turmel, F. (2005). Évaluation de l'implantation et des effets d'un programme de sensibilisation à l'agression indirecte auprès d'élèves du primaire. *Nouveaux cahiers de la recherche en éducation*, 8(2), 17-26.
- *Verreault, M., Pomerleau, A. et Malcuit, G. (2005). Impact de programmes d'activités de lecture sur le développement cognitif et langagier d'enfants âgés de 0 à 5 ans: les programmes ALI. *Éducation et francophonie*, 33(2), 182-206.

- *Verrier, N. (2004). Compétences en langage et perceptions de soi des jeunes enfants : étude longitudinale. *Revue de psychoéducation*, 33(1), 15-39.
- *Vezeau, C., Bouffard, T. et Dubois, V. (2004). Relation entre l'intelligence et les buts d'apprentissage. *Revue des sciences de l'éducation*, 30(1), 9-25.
- *Viau, R., Joly, J. et Bédard, D. (2004). La motivation des étudiants en formation des maîtres à l'égard d'activités pédagogiques innovatrices. *Revue des sciences de l'éducation*, 30(1), 163-176.
- *Vollmeyer, R. et Rheinberg, F. (2004). Influence de la motivation sur l'apprentissage d'un système linéaire. *Revue des sciences de l'éducation*, 30(1), 91-104.
- Williams, R. H. et Zimmerman, D. W. (1989). Statistical power analysis and reliability of measurement. *Journal of general psychology*, 116, 359-369.
- Williams, R. H., Zimmerman, D. W. et Zumbo, B. D. (1995). Impact of measurement error on statistical power : review of an old paradox. *Journal of experimental education*, 63(4), 363-370.

Correspondance

Eric.Yergeau@USherbrooke.ca

Ce texte a été révisé par Sandra Najac.

Texte reçu le : 5 février 2008

Version finale reçue le : 30 janvier 2009

Accepté le : 8 avril 2009