

bonnes pages

On lira ci-après la suite du texte de Gary A. Troia de l'Université de Maryland, Collège Park, États-Unis paru dans *Reading Research Quarterly* (vol.34, n°1, 1er trimestre 1999) qui examine la qualité méthodologique de 39 études ayant entraîné la conscience phonologique chez des enfants.

La première partie (A.L. n°69, mars 2000, pp.61 - 67) était consacrée à ce que l'auteur a appelé les critères de validité interne. Cette deuxième traite des critères de validité externe.

La présence de ce texte dans notre rubrique *Bonnes Pages* n'est évidemment pas étrangère à celle de l'article de Jean Foucambert dans notre précédent numéro (Lire les recherches sur la lecture, A.L. n°69, mars 2000, pp.39-49).

Indépendamment de leur objet, notons que les recherches examinées par A. Troia ont donné lieu à des parutions dans des revues avec Comité de lecture et servent donc, bien souvent, de références scientifiques à d'autres travaux et de cautions à des mises en œuvre ou à des généralisations de pratiques pédagogiques.

INTERVENTIONS DE RECHERCHES SUR LA CONSCIENCE PHONOLOGIQUE : UNE ANALYSE CRITIQUE DE LA METHODOLOGIE EXPERIMENTALE ¹

Gary A. TROIA

♦ LA VALIDITE EXTERNE

Hypothèses de recherche.

Les recherches décrites dans chaque étude s'appuyaient toutes sur une base théorique et/ou empirique justifiable. Les auteurs posaient explicitement ou implicitement leurs hypothèses de recherche après avoir fait une revue de la littérature existante. Ils présentaient ensuite des programmes d'intervention susceptibles de développer ou de préciser les connaissances actuelles sur l'entraînement de la conscience phonologique et, dans certains cas, son impact sur la lecture et/ou l'orthographe.

Description et sélection des participants. La sélection des participants a été correctement décrite dans toutes les études sauf 7 (82%) (Bradley et Bryant, 1983 ; Fox et Routh, 1984 ; Hohn et Ehri, 1983 ; Olofsson et Lundberg, 1983 ; Velutino et Scanlon, 1987 ; Warrick et al., 1993 ; Wise et Olson, 1995). La plupart des investigateurs n'ont pas réalisé une sélection vraiment aléatoire. Ainsi, les enfants ont en général été choisis dans une école ou dans un petit groupe d'écoles d'un seul district. Comme dans la plupart des recherches dans ce domaine et à cause des difficultés qu'ont les chercheurs pour entrer dans les écoles, il est peu probable

¹ Titre original : Phonological awareness intervention research : a critical review of the experimental methodology.

qu'ils aient pu sélectionner des écoles (ou des classes) de manière aléatoire. Il est possible que les écoles sélectionnées aient été désignées pour être sites de recherche par des responsables scolaires et que ces écoles (ou ces classes) aient donc été en quelque sorte particulières, ce qui pourrait représenter un biais systématique de recrutement.

Dans 3 études, les auteurs ont essayé de sélectionner les élèves de façon aléatoire au sein de la population. Olofsson et Lundberg (1983) ont constitué un échantillon de façon aléatoire à partir de toutes les classes maternelles d'une ville. Cependant, un de leurs groupes a été constitué à partir d'une communauté plus petite. Williams (1980) a inscrit dans son étude toutes les classes d'éducation spécialisée de 6 districts scolaires de la ville de New York, même si les districts avaient été choisis par les services administratifs de l'éducation spécialisée. Bentin et Leshem (1993) ont sélectionné des classes maternelles de façon aléatoire, mais uniquement dans des quartiers aisés.

Les caractéristiques démographiques des participants aux études ont été décrites de façon satisfaisante dans 15 (38%) expériences. L'information présentée comprenait le sexe des participants dans les groupes contrôle et les groupes expérimentaux, l'âge moyen des enfants dans chaque groupe, le nombre d'élèves dans chaque niveau, la catégorie socio-économique de l'échantillon et le lieu de l'étude. Cependant, dans la quasi totalité des études, les expérimentateurs n'ont pas fourni d'information sur la composition ethnique des échantillons. La plupart des investigations du corpus ne pourraient pas être reproduites car on ne pourrait pas obtenir des échantillons similaires de participants. De plus, les résultats de la plupart des expériences ne peuvent pas être généralisées car on ne saurait pas bien à quelle population les appliquer.

Sur les 39 études, 26 (67%) ont rapporté à la fois la capacité intellectuelle et la réussite scolaire des participants de l'étude, sept autres (18%) seulement l'un de ces deux points. L'outil le plus fréquemment utilisé pour mesurer le QI était le test « Peabody Picture Vocabulary (révisé) » qui ne donne malheureusement pas une vue d'ensemble des capacités cognitives mais plutôt une mesure de la connaissance de vocabulaire. L'outil de mesure de réussite scolaire le plus utilisé a été le test « Woodcock Reading Mastery (révisé) », même si beaucoup d'autres tests standardisés ainsi que des mesures expérimentales ont aussi été utilisés pour évaluer cette réussite. Les études où n'étaient évaluées ni les capacités intellectuelles ni la réussite scolaire ne comprenaient que des participants provenant de classes ordinaires. Cependant, sans

cette information, la véracité de l'hypothèse de normalité est compromise.

Aucun des auteurs n'a présenté suffisamment d'information concernant les élèves en éducation spécialisée. Bien que les auteurs des études incluant des enfants ayant des besoins spécifiques aient indiqué que les participants recevaient une éducation spécialisée, ils n'ont pas précisé le type et l'intensité de ces activités ou depuis combien de temps ils suivaient cette éducation spécialisée (mais voir Korkman et Peltomaa, 1993, pour une tentative louable). Aucune des études où les participants étaient des élèves connaissant une réussite normale n'a fourni une information adéquate sur l'histoire ou les expériences précédentes de ces élèves en éducation spécialisée, bien que Hatcher et al. (1994) essayèrent remarquablement de le faire.

Trois des sept études incluant des enfants avec des besoins spécifiques n'ont pas donné les critères explicites permettant de classifier et de décrire les handicaps. Cette information est nécessaire pour pouvoir sélectionner des échantillons homogènes d'enfants handicapés, puisque les échantillons représentant les écoles sont souvent hétérogènes. Williams (1980) a simplement déclaré que ses élèves étaient inscrits dans des classes pour enfants avec difficultés d'apprentissage. Cependant, la mesure du QI d'un sous-échantillon de ses participants suggère que ses enfants se situaient en dessous de la moyenne. De plus, chaque élève n'a fait ni le même test de QI ni le même test de réussite en lecture, probablement parce que les tests n'étaient pas conduits par l'investigateur mais par les responsables des différents districts scolaires. Ainsi, les différences dans la validité et la fiabilité des tests ont pu compromettre l'identification d'un échantillon homogène. Néanmoins, il n'y avait pas, dans ces mesures, de différence statistiquement significative entre les groupes expérimentaux et les groupes contrôle.

O'Connor et al. (1993) ont indiqué que leurs élèves connaissaient différents retards de développement, dont des difficultés du langage, des retards mentaux, des troubles émotionnels et des handicaps physiques. Heureusement, ils ont fait passer un test de QI à tous leurs participants et n'ont pas trouvé de différences statistiquement significatives entre les groupes. Dans une étude ultérieure, O'Connor et ses collègues (1995) ont déclaré que leur échantillon comprenait des élèves qui suivaient une éducation spécialisée ou qui allaient y être envoyés, ainsi que des élèves sans handicaps particuliers. Aucune autre information sur les enfants avec handicaps n'a été donnée. Bien que les résultats individuels de QI des élèves de l'échantillon variaient

beaucoup, il ne semblait pas y avoir de grandes différences entre les moyennes des groupes.

Les mesures de transfert et de maintien des compétences. Le simple fait de dire aux élèves qu'on attend d'eux qu'ils utilisent leurs connaissances dans des contextes non familiers est souvent favorable au transfert de l'apprentissage (ex. Derry, 1990 ; Ellis, Lenz et Sabornie, 1987), ce qui n'a été fait que dans 5 études (13%). La plupart des investigateurs n'a pas fait cet effort, soit parce qu'ils n'avaient pas conscience de ce principe, soit parce qu'ils avaient choisi de l'ignorer. Presque tous les investigateurs (90%) ont évalué le transfert des compétences acquises en matière de conscience phonologique à d'autres types d'activités. Par exemple, Lundberg et al. (1988) ont conduit des enquêtes sur la généralisation 3 et 10 mois après que les interventions relatives à la conscience phonologique aient eu lieu. Ces enquêtes comprenaient un test de reconnaissance de mots fait en groupe, dans lequel les élèves devaient sélectionner, parmi quatre choix, une image sur laquelle un mot était écrit, un test d'orthographe à l'oral et un test mathématique en groupe. Hohn et Ehri (1983) ont fait faire un test de décodage de non-mots, immédiatement avant et après un entraînement des compétences d'analyse de phonèmes. De plus, un exercice d'effacement de phonèmes, avec des sons connus et inconnus dans des non-mots, a été faite après le test. L'exercice d'effacement de phonèmes n'avait pas été inclus dans le programme d'intervention.

Dans presque la moitié (46%) des études, les chercheurs ont clairement évalué le transfert des compétences relatives à la conscience phonologique par de nouveaux stimuli. Par exemple, Slocum et ses collègues (1993) ont fait passer des pré-tests, des tests intermédiaires et des post-tests de segmentation et d'enchaînement de phonèmes comportant 10 items qui n'avaient pas été étudiés pendant la formation. Ball et Blachman (1991) ont fait faire un pré-test et un post-test de segmentation de phonèmes comprenant : 12 items qui avaient fait l'objet d'un entraînement, 12 items proches de l'entraînement (phonèmes non entraînés appariés à des phonèmes entraînés, selon des caractéristiques articulatoires) et 10 items éloignés de l'entraînement (phonèmes non entraînés, différents de tous ceux traités dans l'intervention).

Des enquêtes de maintien ont été conduites dans moins du tiers (31%) des investigations. La durée des effets du traitement a été évaluée entre 10 semaines (Weiner, 1994) et 3 ans (Byrne et Fielding-Barnsley, 1991, 1993, 1995) après la fin

de l'entraînement. Bien qu'aucun des investigateurs n'ait formellement évalué la validité sociale de leurs interventions (une source de variation potentielle pour le maintien des effets observés), certains ont fait remarquer que les participants semblaient aimer les activités d'entraînement (ex. Foster et al., 1994 ; Olofsson et Lundberg, 1983) ou que la plupart des enseignants pensaient que les interventions expérimentales étaient bénéfiques aux élèves (ex. Williams, 1980).

En résumé, toutes les études reposent sur des principes théoriques et/ou des recherches antérieures solides. La plupart des expérimentateurs ont décrit de manière adéquate la sélection de leurs participants dans la population, bien qu'aucun d'eux n'ait fait une sélection vraiment aléatoire. Les investigateurs ont décrit de façon adéquate la composition démographique de leurs échantillons dans seulement un peu plus d'un tiers des études. Peu de chercheurs ont donné des informations sur les origines ethniques ou socio-économiques de leurs élèves.

La plupart des études ont évalué les capacités intellectuelles ou la réussite scolaire mais aucune n'a présenté d'informations détaillées sur le passage des participants en éducation spécialisée. Parmi les études incluant des enfants ayant des besoins spécifiques, la moitié seulement ont adopté des critères explicites de classification des handicaps. Presque toutes les études mesurent le transfert des compétences mais moins de la moitié étudient clairement la généralisation du stimulus. La durée des effets du traitement a été évaluée dans seulement un tiers des investigations.

Huit études (21%) satisfont au moins les trois quart des critères de validité externe applicables (Blachman et al., 1994 ; Brady et al., 1994 ; Byrne et Fielding-Barnsley, 1991, 1993, 1995 ; Cunningham, 1990 ; Hatcher et al., 1994 ; Kennedy et Backman, 1993 ; Slocum et al., 1993 ; Uhry et Shepherd, 1993). Parmi elles, 3 ne respectaient pas un critère classé comme essentiel pour pouvoir généraliser les conclusions de l'étude.

Dix autres études (26%) satisfont au moins les deux tiers des critères de validité externe, et 4 d'entre elles présentent au moins un défaut majeur.

Le motif est dans le tapis.

Henry James.
à propos d'un texte.

Discussion

Les défauts méthodologiques les plus sérieux observés dans de nombreuses études sur la conscience phonologique ont été :

- a) répartition non aléatoire des participants dans les groupes expérimentaux ;
- b) non contrôle de l'effet Hawthorne ;
- c) réalisation fidèle de l'expérience insuffisamment ou non contrôlé ;
- d) mauvaise sensibilité de la mesure ;
- e) échantillons décrits de manière inadéquate.

D'autres défauts méthodologiques, bien qu'ils ne soient pas assez graves pour qu'on estime que les conclusions ne sont pas interprétables ou pas valides, ont été observés dans la majorité des études :

- a) différences dans l'exposition aux différents matériels expérimentaux entre les groupes ;
- b) confusion enseignants / conditions ;
- c) enseignement non basé sur un critère ;
- d) données sur la fiabilité des mesures dépendantes inexistantes ou non communiquées ;
- e) utilisation d'unités d'analyse non appropriées ;
- f) non prise en compte des questions de transfert de stimulus ;
- g) non évaluation des effets de maintien.

À cause de ces faiblesses, un grand nombre des recherches dans ce domaine doit être considéré avec prudence.

Des résultats semblables à ceux présentés dans cette étude ont été obtenus dans une analyse méthodologique de recherches sur l'enseignement de stratégies de compréhension en lecture, conduite par Lysynchuk, Pressley, d'Ailly, Smith et Cake (1989) avec des enfants de la maternelle au huitième grade (équivalant approximativement en France à la 4ème, NDT), ainsi que dans une étude similaire sur la recherche en enseignement de stratégies de lecture conduite par Ridgeway, Dinston et Qian (1993) dans des écoles

secondaires. Les auteurs de ces études ont souligné les problèmes liés à la validité interne et externe des recherches évaluées, de par :

- a) un défaut de randomisation ;
- b) des effets Hawthorne possibles ;
- c) une inégalité entre les taux de perte ;
- d) des différences dans l'exposition aux matériels d'entraînement entre les participants des groupes expérimentaux et les participants des groupes contrôle ;
- e) des temps d'intervention inégaux selon les conditions d'entraînement ;
- f) une confusion enseignant / condition ;
- g) un défaut de contrôle des manipulations ;
- h) une mauvaise sensibilité des mesures et un manque de données concernant la fiabilité des mesures dépendantes ;
- i) une utilisation d'unités d'analyse inadéquates ;
- j) une absence de tests de généralisation et de maintien.

Apparemment, les chercheurs ont continué à ignorer des principes méthodologiques importants dans leurs recherches sur l'enseignement, malgré de telles études. Le principe des comités de lecture est en partie responsable de cet état de fait. Le premier objectif d'un comité de lecture devrait être la sélection de recherches parfaites d'un point de vue méthodologique et pertinentes d'un point de vue empirique. La prolifération de recherches mal conduites compromet gravement l'objectif premier : transposer les recherches dans les pratiques en classe.

Contrairement aux études mentionnées ci-dessus, nous avons pondéré les critères d'évaluation en fonction de leur contribution à la rigueur méthodologique. Cela nous a permis de classer chaque étude en fonction de sa solidité méthodologique par rapport aux autres études du corpus. Par conséquent, nous pouvons, en toute confiance, en tirer certaines conclusions.

Sept études du corpus (18%) satisfaisaient au moins les deux tiers des critères d'évaluation de la validité interne et externe (voir tableau 4), bien qu'elles comprenaient toutes entre 1 et 3 défauts majeurs (ex. : pas de preuve suffisante quant à la fidélité du traitement). En considérant ces 7 études comme les meilleurs exemples disponibles, on peut conclure que l'entraînement métaphonologique peut améliorer à la fois les compétences relatives à la conscience phonologique analytique et synthétique et l'acquisition de la lecture et de l'écriture (au moins les compétences de bases

La fiction n'est que découverte...

Les essais en savent trop.

Cynthia Ozick.

Art & Ardor

en lecture, voir Uhry et Shepherd, 1993), avec un enseignement individuel ou en petit groupe ne dépassant pas 2 mois (les séances durant souvent 15-20 minutes et ayant lieu en général deux fois par semaine).

Malheureusement, aucune de ces études n'a évalué les effets d'une intervention dans une classe. Nous n'avons donc pas la preuve que les programmes d'entraînement de la conscience phonologique sont écologiquement valides et efficaces dans un environnement scolaire. Il est possible que de tels programmes soient difficilement applicables ou trop complexes pour être mis en œuvre en classe et que les effets positifs observés par les chercheurs soient compromis dans ces conditions. De plus, seules 2 des 7 études (Hatcher et al., 1994 ; Slocum et al., 1993) ont contrôlé la fidélité d'application du traitement, mais aucune des deux n'a apporté la preuve que cette application s'était effectuée de manière fiable. Sans cette information, on ne peut pas savoir jusqu'à quel point ces programmes peuvent supporter des modifications ou des variations, dans leurs applications, sans que l'efficacité du traitement en soit diminué.

Une seule étude (Byrne et Fielding-Barnsley, 1991, 1993, 1995) a montré que les effets du traitement peuvent se maintenir dans le temps (pendant 3 ans), au moins pour ceux des élèves qui, au début, avaient bien répondu au traitement. La durée de ces effets doit être démontrée de façon plus convaincante : si les bénéfices dus à l'intervention sont de courte durée, les fonds accordés aux programmes d'entraînement de la conscience phonologique sont mal employés. De plus, Byrne et Fielding-Barnsley (1991, 1993, 1995) ainsi que Torgesen et ses collègues (1992) ont montré que tous les enfants ne tirent pas bénéfice d'un entraînement de la conscience phonologique. Après l'évaluation de leur conscience phonologique ou de leur réussite en lecture, certains élèves de leurs échantillons n'ont pas fait de progrès notables.

Il est possible que des programmes d'entraînement plus élaborés, plus longs et/ou plus intensifs soient nécessaires pour que ces enfants progressent en lecture et en écriture, ou qu'une approche complètement différente soit requise. Malheureusement, ces études proposent peu d'informations sur l'efficacité comparée de différents types d'interventions relatives à la conscience phonologique ou sur les comparaisons possibles entre ce type d'entraînement et d'autres formes d'intervention. Enfin, les chercheurs de ces 7 études ont étudié les effets de l'entraînement sur des enfants de 4 à 7 ans connaissant une réussite normale, à

un âge où les compétences en lecture et en écriture progressent rapidement. Par conséquent, nous n'avons pas de données fiables sur l'efficacité de l'entraînement métaphonologique chez des élèves plus âgés ou chez des élèves avec des difficultés, même si certains investigateurs aient inclus dans leur recherche des élèves exposés à rencontrer plus tard des difficultés en lecture (Castle et al., 1994 ; Slocum et al., 1993 ; Torgesen et al., 1992).

Nous devons souligner deux limites à cette critique méthodologique. Premièrement, seules les études parues dans des journaux à comité de lecture ont été incluses, ce qui introduit un biais dans le corpus. Comme nous l'avons déjà fait remarquer, il n'est pas étonnant que des résultats positifs aient été obtenus dans toutes les investigations : les résultats nuls ne sont souvent pas publiés dans les journaux (Atkinson, Furlong et Wampold, 1982). Des méta-analyses comme celles conduites par Hattie et ses collègues (Hansford et Hattie, 1982 ; Hattie, Biggs et Purdie, 1996) ont montré qu'il y a plus d'interventions efficaces et de corrélation entre les variables dans les journaux que dans les mémoires ou les conférences. Plus important peut-être pour l'objet de notre étude, la rigueur méthodologique des études sélectionnées n'est peut-être pas représentative de l'ensemble des recherches sur les entraînements de la conscience phonologique.

Deuxièmement, notre étude a opposé le contrôle expérimental à la validité externe. Cela se lit facilement dans les données présentées dans le tableau récapitulatif. En général, les études qui satisfont à une plus grande proportion de critères de validité interne, satisfont en même temps une moins grande proportion de critères de validité externe. Ridgeway et ses collègues (1993) ont rapporté des conclusions similaires dans leur revue de méthodologie. Ils ont trouvé une relation inverse, significative d'un point de vue statistique ($r = .45$) entre la validité interne et la validité écologique (c'est à dire la possibilité de généralisation à d'autres environnements) dans les études relatives à l'enseignement de stratégies d'apprentissage.

Les études sur l'entraînement de la conscience phonologique les mieux classées, satisfaisant à une majorité de critères d'évaluation, étaient les moins représentatives d'un enseignement en classe : elles avaient été effectuées par des enseignants spécialement formés ou par des chercheurs, à l'extérieur des classes, avec des petits groupes d'élèves ou individuellement, sur des enfants issus d'échantillons de population aux caractéristiques précises. En raison de la difficulté extrême, voire de l'impossibilité de contrôler tout

ce qui menace la validité interne en situation scolaire, ces recherches doivent trouver un compromis entre le contrôle expérimental et la possibilité de généraliser (Bracht et Glass, 1968 ; Ridgeway et al., 1993 ; Snow, 1974).

Néanmoins, ce sont ces recherches dans des classes qui sont le plus nécessaire pour améliorer les pratiques d'enseignement (Howe et Eisenhart, 1990 ; Snow, 1974). Bien que les chercheurs ne doivent pas se dégager de leur responsabilité dans l'emploi de méthodes de recherche solides, la communauté éducative doit être prête à accepter des compromis et innovations dans la méthodologie expérimentale et des paradigmes d'investigation différents, afin que des traitements fiables au sein de différents environnements soient disponibles sur le terrain. Cependant, nous pouvons faire, à partir des résultats de cette étude, un certain nombre de recommandations qui pourront aider les chercheurs en interventions à résoudre les dilemmes auxquels ils se trouvent confrontés lorsqu'ils essaient de conduire des études à la fois solides d'un point de vue empirique et pertinentes sur le plan éducatif.

Quand l'affectation aléatoire des participants n'est pas possible (comme c'est en général le cas dans des recherches en classe), il est fondamental que les groupes entiers d'élèves soient affectés de manière aléatoire aux conditions de traitement, afin d'avoir le droit de parler de causalité. Les chercheurs doivent s'efforcer de contrôler les effets Hawthorne en faisant suivre aux groupes contrôles des interventions pédagogiques concurrentes mais sans effet sur le point étudié. Les chercheurs peuvent, si cela s'avère nécessaire, faire faire des exercices supplémentaires aux participants contrôle, afin que les temps d'enseignement soient équilibrés entre les différentes conditions. Si les descriptions approfondies de chaque condition d'entraînement sont trop longues pour être incluses dans le corps du rapport de recherche, elles peuvent y être annexées. Lorsque cela est possible, les enseignants doivent avoir la charge de l'enseignement dans chacune des conditions expérimentales, afin d'éviter la confusion enseignant/condition. Si cela n'est pas possible, les enseignants doivent alors être affectés de manière aléatoire aux différentes conditions de traitement.

Quand un nombre important d'élèves faisant partie d'une certaine condition expérimentale se retire de l'étude, les investigateurs doivent évaluer la probabilité que ces élèves soient différents de ceux qui restent, en examinant leurs résultats aux tests standards et aux mesures d'évaluation effectuées au cours du pré-testage. Ces recommandations

sont autant de moyens pour les chercheurs d'améliorer le plan général des études sur l'entraînement, surtout celles conduites dans des classes.

On pourra observer des séances d'enseignement et utiliser des listes de contrôle afin d'évaluer la validité de la mise en œuvre du traitement : ceci est un point important sur le plan des mesures dans les études d'entraînements. De plus, les chercheurs devraient effectuer d'avantage d'études pilotes pour améliorer la fiabilité et la sensibilité des mesures dépendantes. Si l'enseignement est effectué dans des groupes d'élèves, la moyenne du groupe doit être l'unité d'analyse dans les procédures statistiques, à moins qu'une justification du non-respect de cette convention soit apportée ou que d'autres procédures statistiques soient utilisées. Chacun doit s'efforcer au maximum de fournir toutes les statistiques descriptives utiles (ex. moyennes et déviations standards) et de calculer les tailles des effets, afin de faciliter l'interprétation des résultats.

Enfin, pour améliorer les possibilités de généralisation des conclusions des recherches en interventions, les chercheurs devraient utiliser des tableaux présentant les caractéristiques principales des enfants de l'échantillon sélectionné (dont les données sur les capacités et la réussite au test). Des mesures de transfert (à la fois sur le stimulus et sur la tâche) devraient être faites afin d'évaluer l'importance de l'effet de l'entraînement, et des enquêtes de maintien devraient être conduites pour évaluer la durée des effets du traitement.

La diversité des programmes d'entraînement relatifs à la conscience phonologique développés par les chercheurs et évalués dans la littérature spécialisée existe et renforce notre compréhension de cet aspect important de la réussite en lecture et en écriture, tandis que les limites liées à la méthodologie utilisée dans la plupart de ces recherches ne font que limiter cette compréhension. Il y a dix ans, Lysynchuk et ses collègues (1989) ont demandé à la communauté de chercheurs de mieux s'auto-contrôler afin que les recherches vicieuses (VIDE DE FORME) ne puissent pas influencer les politiques et les pratiques éducatives. Il reste beaucoup à faire pour améliorer la qualité des recherches en éducation dans ce domaine ainsi que dans d'autres, afin qu'elles puissent résister à l'examen minutieux de ceux qui sont le plus touchés par les efforts des chercheurs en interventions : les enseignants.

Traduction : *Magali FRANÇOIS, Léna COÏC,
Denis FOUCAMBERT*

	% de critères de validité interne satisfaits	nombre de défauts cruciaux	% de critères de validité externe satisfaits	nombre de défauts cruciaux	note pondérée	classement
- Ball & Blachman, 1988,1991	65	2	60	1	23	11
- Bentin & Leshem, 1993	47	1	67	0	23	10
- Blachman, Ball, Black & Tangel, 1994	47	4	73	0	24	13
- Bradley & Bryant, 1983, 1985	47	4	47	2	35	32
- Brady, Fowler, Stone & Winbury, 1994	27	4	73	0	29	21
- Byrne & Fielding-Barnsley, 1991, 1993, 1995	65	2	73	1	19	-
- Cary & Verhaeghe, 1994 (Expérience 1)	20	4	67	0	33	30
- Cary & Verhaeghe, 1994 (Expérience 2)	35	3	60	1	33	29
- Castle, Riach & Nicholson, 1994 (Expérience 1)	59	1	67	0	20	8
- Castle et al., 1994 (Expérience 2)	65	1	67	0	19	5
- Content, Morais, Alegria & Bertelson, 1982	41	3	67	2	29	22
- Cunningham, 1990	53	2	73	0	21	9
- Foster, Erickson, Foster, Brinkman & Torgesen, 1994 (Expérience 1)	80	2	60	1	19	6
- Foster et al., 1994 (Expérience 2)	73	3	40	2	28	18
- Fox & Routh, 1984	53	4	53	3	32	28
- Haddock, 1976	53	3	47	3	32	27
- Hatcher, Hulme & Ellis, 1994	67	1	73	1	17	3
- Hohn & Ehri, 1983	60	2	53	2	25	14
- Kennedy & Backman, 1993	33	3	81	0	23	12
- Korkman & Peltomaa, 1993	40	4	63	1	31	24
- Kozminsky & Kozminsky, 1995	47	3	53	2	33	31
- Lie, 1991	47	3	67	1	26	16
- Lundberg, Frost & Peterson, 1988	27	4	67	1	32	26
- McGuinness, McGuinness & Donohue, 1995	40	3	60	1	29	21
- O'Connor, Jenkins, Leicester & Slocum, 1993	53	2	56	2	29	20
- O'Connor, Jenkins & Slocum, 1995	59	2	50	3	30	23
- Olofsson & Lundberg, 1983, 1985	31	4	60	3	35	33
- Rosner, 1974	40	4	40	3	40	35
- Slocum, O'Connor & Jenkins, 1993	71	1	73	1	15	2
- Tangel & Blachman, 1992	47	4	67	0	26	15
- Torgesen & Davis, 1996	40	3	60	1	29	21
- Torgesen, Morgan & Davis, 1992	82	2	67	0	14	1
- Tornéus, 1984	40	3	40	4	39	34
- Uhry & Shepherd, 1993	65	2	73	0	18	4
- Vellutino & Scanlon, 1987	53	3	56	2	32	25
- Warrick, Rubin & Rowe-Walsh, 1993	27	5	50	3	41	36
- Weiner, 1994	47	3	67	2	26	17
- Williams, 1980	40	4	50	3	35	33
- Wise & Olson, 1995	53	2	60	2	29	19