

CONFÉRENCE DE COMPARAISONS INTERNATIONALES

GOUVERNANCE DES POLITIQUES ÉDUCATIVES



RAPPORT THÉMATIQUE

**Efficacité des interventions conduites dans les classes :
la nécessité de l'évaluation de leur implémentation**

#CCI_GOUVERNANCE

MARS 2022

EFFICACITÉ DES INTERVENTIONS CONDUITES DANS LES CLASSES : LA NÉCESSITÉ DE L'ÉVALUATION DE LEUR IMPLÉMENTATION

Édouard GENTAZ

Université de Genève, Suisse & CNRS, France,

Sylvie RICHARD

**Haute École Pédagogique du Valais / HEPVS & Université de
Genève, Suisse**

Mars 2022

le cnam
Cnesco

Centre national d'étude des systèmes scolaires

Pour citer ce document, merci d'utiliser la référence suivante :
Gentaz, E. & Richard, S. (2022). *Efficacité des interventions conduites dans les classes : la nécessité de l'évaluation de leur implémentation*. Paris : Cnesco-Cnam.

Ce texte s'inscrit dans une série de rapports publiés par le Centre national d'étude des systèmes scolaires (Cnesco) sur la thématique : **Gouvernance des politiques éducatives**.

Les opinions et arguments exprimés n'engagent que les auteurs du rapport.

Auteurs :
Édouard Gentaz
Sylvie Richard

Disponible sur le site du Cnesco : <http://www.cnesco.fr>
Publié en mars 2022.
Centre national d'étude des systèmes scolaires
41 rue Gay-Lussac 75005 Paris

Contact : cnesco@lecnam.net - 06 98 51 82 75

Sommaire

Introduction : recherches scientifiques et éducation.....	5
I. Les preuves scientifiques en éducation	7
A. La démarche de « Médecine fondée sur des preuves »	7
B. Les deux démarches scientifiques dans le domaine de l'éducation	9
C. Un exemple des difficultés à généraliser à grande échelle les résultats de la recherche sur l'apprentissage de la lecture en France.....	10
1. Les objectifs de l'intervention.....	10
2. Considérations méthodologiques et résultats.....	11
3. Limites.....	12
4. Conclusions	13
II. Les apports des sciences de l'implémentation des interventions	14
A. Facteurs influençant le processus d'implémentation dans le domaine de la santé	14
B. Définition et composantes liées à l'implémentation	16
C. La mesure de la fidélité d'implémentation	17
D. Exemple d'une démarche de conception d'outils didactiques en France	19
E. Exemple d'une recherche interventionnelle visant à développer les compétences socio-émotionnelles chez des enfants de 5-6 ans : processus d'implémentation d'un programme en contexte scolaire dans le Canton du Valais en Suisse.	20
Conclusion, limites et perspectives	24
A. La démarche collaborative	24
B. L'appropriation d'un outil innovant	24
C. L'accompagnement de l'enseignant	25
D. La triade : « vouloir » / « pouvoir » / « savoir »	25
Références bibliographiques.....	27

Introduction : recherches scientifiques et éducation

Un grand nombre d'institutions à travers le monde soulignent l'importance de la prise en compte des données scientifiques dans la politique éducative, suggérant l'avènement d'une politique éducative fondée sur des preuves. Cette approche est incarnée depuis de nombreuses années par les chercheurs en sciences cognitives appliquées à l'éducation (Dessus & Gentaz, 2006 ; Gentaz & Dessus, 2004). Comment favoriser alors une dynamique de transfert entre, d'une part, les chercheurs et leurs preuves scientifiques et, d'autre part, les enseignants et leurs savoirs empiriques ? Plusieurs leviers sont disponibles.

Un premier levier concerne l'état de la recherche à communiquer aux enseignants : ce dernier doit faire l'objet d'un consensus clairement établi par une expertise collective regroupant l'ensemble des disciplines scientifiques appliquées à l'école. Les rapports d'expertise produits par l'Inserm en France (par exemple celui publié en 2007 sur la dyslexie) ou par le *National Reading Panel* (2000) aux États-Unis sur l'apprentissage de la lecture sont des exemples très pertinents. Les résultats à diffuser doivent principalement se fonder sur les conclusions de méta-analyses et non sur quelques études isolées, même lorsque celles-ci sont publiées dans de grandes revues scientifiques et présentent des résultats potentiellement prometteurs. Rappelons qu'une méta-analyse est une méthode scientifique systématique qui combine les résultats d'une série importante d'études indépendantes sur un sujet donné. Afin de comparer et de synthétiser les résultats statistiquement significatifs issus de différentes études, les méta-analyses mesurent généralement l'efficacité de méthode ou intervention grâce à une mesure statistique, appelée taille d'effet, qui fournit un indice quantitatif de la force de l'intervention testée (Laroche, 2015). En d'autres termes, les connaissances à diffuser auprès des enseignants doivent apporter des données claires et transparentes sur « le degré de certitude » disponible sur l'efficacité de telle ou telle pratique, comme le fait le site de la fondation « *Education Endowment Foundation* (EEF) » ou la graduation en 3 niveaux « *Strong evidence, moderate evidence, promising evidence* » préconisée par l'« *Every Student Succeeds Act (ESSA)* » aux États-Unis (Slavin, 2020).

Un second levier concerne la confiance des lecteurs envers la crédibilité des informations à caractère scientifique publiées par la revue ou le guide dépend en partie de la transparence de la gestion des conflits d'intérêts pouvant intervenir à tous les niveaux du déroulement d'une étude scientifique, de sa planification à sa publication (Gentaz, 2019). Ajoutons que la suspicion de conflits d'intérêts (réels ou non) est tout aussi importante que les conflits d'intérêts eux-mêmes car elle engendre des effets délétères similaires. Rappelons qu'un conflit d'intérêt existe quand un auteur et/ou un co-auteur a des relations financières ou personnelles avec d'autres personnes ou organisations qui sont susceptibles d'influencer ses jugements professionnels concernant une valeur essentielle (bien du patient, intégrité de la recherche...). Les conflits d'intérêt peuvent être des intérêts financiers, professionnels ou encore relationnels. Il est donc important que tous les auteurs de publications déclarent toutes les relations qui peuvent être considérées comme de potentiels conflits d'intérêt en lien avec le texte publié. Si cette pratique est relativement bien présente et exigée maintenant dans les sciences biologiques, médicales ou agronomiques compte tenu des enjeux financiers et de santé publique, cette pratique est peu usitée dans les sciences cognitives et humaines. Or, ces sciences, quand elles ont des dimensions appliquées ou interventionnelles peuvent aborder des questions ou des thèmes ayant une forte implication sociétale.

Cependant, comme l'a reconnu le psychologue William James il y a plus d'un siècle, les descriptions psychologiques ne peuvent se transposer directement en prescriptions éducatives spécifiques (Mayer, 1992) : la raison principale en est qu'on ne peut pas déduire directement de résultats de recherches, spécifiques à une situation, des procédures détaillées, applicables à toute situation, pour tout public, même s'il existe aussi une perméabilité entre pédagogie et science. Nous allons voir que même si des résultats de recherche produisent des preuves scientifiques fiables, leur mise en application dans les classes demande encore un effort très important de recherche devant mobiliser conjointement les acteurs du monde de la recherche et ceux du monde éducatif. Les recherches sur l'implémentation décrivent les conditions qui favorisent l'efficacité des interventions à grande échelle conduites en classe par les enseignants.

Les termes « interventions, programmes, outils, méthode » en contexte scolaire sont utilisés de manière interchangeable dans le cadre de cette contribution. Ils englobent à la fois des pédagogies mises en œuvre au quotidien en classe, comme c'est le cas du programme « *Tools of the Mind* » (Bodrova & Leong, 2001) ou celle de la pédagogie Montessori (Gentaz, Richard, & Rivier, 2021) ou la mise en œuvre d'activités d'une durée plus ou moins longue centrée sur des apprentissages scolaires ou transversaux à différents niveaux scolaires. Il existe de nombreuses interventions en classe qui ont été évaluées en France depuis plusieurs décennies : plusieurs ont été proposées à des élèves de 5-6 ans dans le but de favoriser l'apprentissage de la lecture (Bara, Gentaz & Colé, 2007 ; Bara, Gentaz, Cole & Sprenger-Charolles, 2004 ; Gentaz, Colé & Bara, 2003 ; Labat, Ecalte, Baldy & Magnan, 2014 ; Labat, Magnan & Ecalte, 2011), de l'écriture (Bara & Gentaz, 2011), de la géométrie (Kalénine, Pinet & Gentaz, 2011 ; Pinet & Gentaz, 2008). D'autres ont été proposées à des élèves plus âgés dans le but de favoriser des compétences transversales, comme les compétences émotionnelles chez les élèves de 6 à 12 ans (Theurel & Gentaz, 2015) ou métacognitives chez des élèves de 16 ans (Malsert, Theurel & Gentaz, 2013). Deux exemples de recherches interventionnelles seront présentés en détail plus loin, une visant à développer les compétences en lecture chez des élèves de 6-7 ans en France (Gentaz *et al.*, 2013) et une autre visant à développer les compétences socio-émotionnelles chez des élèves de 5-6 ans en Suisse (Richard, Clerc-Georgy, Baud-Bovy & Gentaz, 2021).

I. Les preuves scientifiques en éducation

A. La démarche de « Médecine fondée sur des preuves »

Le principe de la démarche de « Médecine fondée sur des preuves » (*Evidence-Based Medicine - EBM*) est basé sur le fait que toutes les preuves ne se valent pas (elles sont donc à hiérarchiser) et que la pratique des professionnels doit se fonder sur les meilleures preuves disponibles (Gentaz, 2020, in press). Ce concept a été développé à partir des années 1980-1990 à l'université de médecine Mc Master au Canada (Haynes, Straus, Glasziou & Richardson, 2007). L'objectif initial de l'EBM était de former les médecins-cliniciens à la compréhension et à l'utilisation des résultats issus de la littérature scientifique (en constante augmentation), afin de donner à un patient les meilleurs soins possibles selon les meilleures données scientifiques disponibles et ses caractéristiques. Il est classique de décrire quatre phases pour résoudre un problème clinique concernant un patient donné : 1) identifier le problème médical ; 2) rechercher les articles pertinents ; 3) évaluer la validité et la pertinence des résultats trouvés et 4) intégrer les résultats retenus pour le patient. Il est à noter que les revues systématiques en médecine et santé appelées Cochrane sont l'un des maillons fondamentaux de la démarche EBM¹. Il existe une organisation internationale indépendante similaire pour de nombreux domaines des sciences sociales, appelée Campbell Collaboration², intégrant l'éducation. Malheureusement, l'impact des travaux publiés dans *Campbell Systematic Reviews* est beaucoup plus faible que ceux publiés dans *Cochrane Database of Systematic Reviews*.

Cette démarche avec l'« *Evidence-Based Practice* » a été étendue à d'autres domaines : ceux liés à la médecine (orthophoniste-logopédiste, neuropsychologue, etc.) mais aussi ceux liés à l'éducation et à l'enseignement (éducateur, enseignant, etc.). Cette démarche est également de plus en plus utilisée dans le monde politique pour rationaliser les politiques publiques (« *Evidence-Based Policy* »).

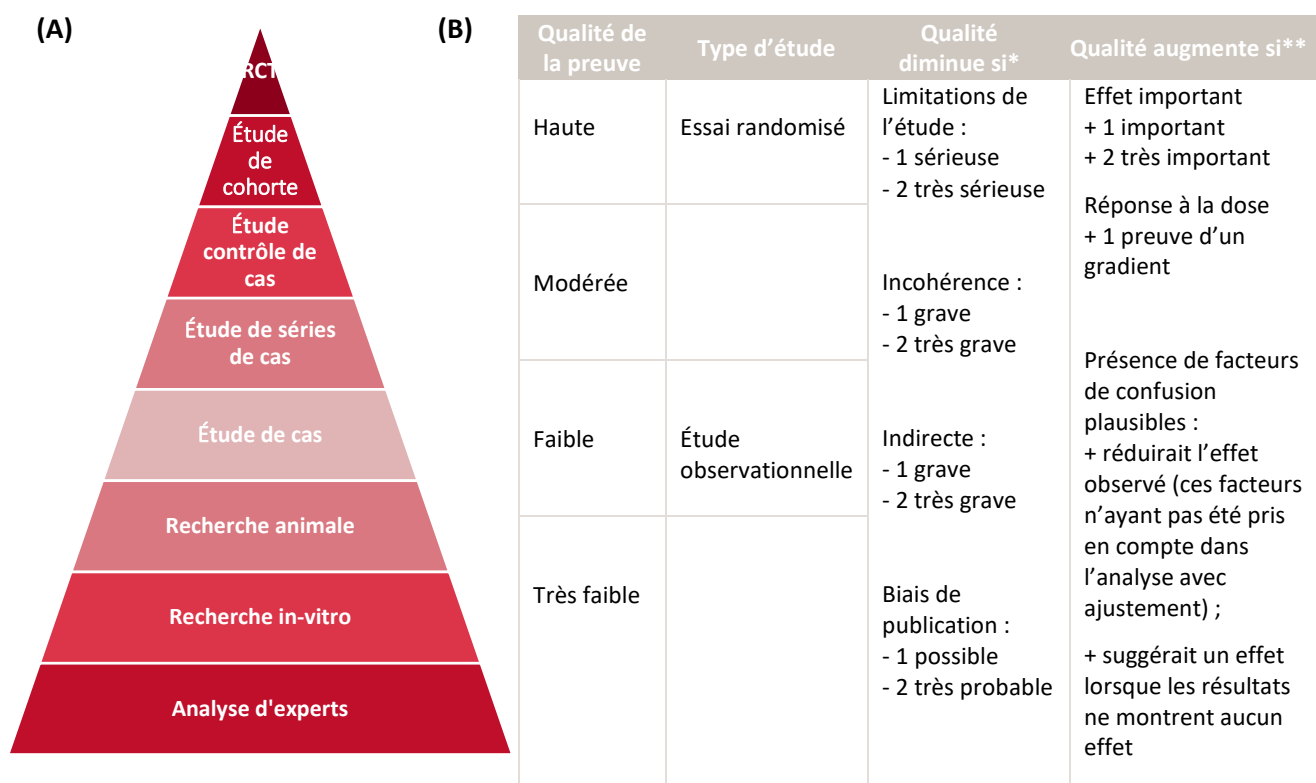
Depuis 30 ans, la démarche EBM est, en soi, un domaine de recherche et comme elle fait l'objet de critiques positives et négatives, sa conception et son application évoluent régulièrement. Du fait de sa diffusion dans d'autres domaines que la médecine, il semble important de prendre en compte les résultats de ces recherches en médecine afin d'éviter de reproduire les mêmes erreurs dans les autres domaines, par exemple celui de la recherche en éducation et des recommandations pédagogiques. Dans une revue critique publiée dans la revue médicale *The Lancet*, Djubegovic et Guyatt (2017) analysent une avancée majeure des recherches particulièrement pertinente dans le cadre de ce rapport. L'EBM reconnaît maintenant les limites des seules preuves scientifiques et souligne la nécessité de combiner l'évaluation critique des preuves avec les valeurs et les préférences des patients par le biais d'une prise de décision partagée. Les auteurs proposent de comparer la conception ancienne pyramidale (*Figure 1A*) à la nouvelle conception de classification de la qualité des preuves appelée *GRADE – Grading of Recommendations Assessment, Development and Evaluation* (*figure 1B*). Il est important de souligner que l'EBM ne se limite pas aux résultats issus des études randomisées en double aveugle (RCT), qui doivent aussi être évaluées qualitativement par un comité de pairs (autres chercheurs) et peuvent également présenter des limites. L'EBM peut en effet aussi intégrer des études de cohortes, de séries de cas ou des observations (pouvant elles-mêmes présenter des limites). De

¹ <https://ccf.cochrane.org/revues-cochrane>

² <https://www.campbellcollaboration.org/>

plus, l'EBM prévoit la prise en compte de l'ensemble des données disponibles (de tous les niveaux) et non uniquement une sélection biaisée d'études dans une revue systématique. Enfin, la bonne pratique de l'EBM du médecin pour prendre une décision efficace implique la prise en compte non seulement de ces niveaux de preuves mais aussi l'environnement et le contexte du patient avec ses préférences et ses valeurs.

Figure 1 : Hiérarchie des preuves : EBM traditionnelle *versus* GRADE. Comparaison de la hiérarchie des preuves de l'EBM traditionnelle (1991-2004) avec la classification GRADE de la qualité des preuves en 4 niveaux (de 2004 à aujourd'hui) (adapté de Djubegovic, Guyatt & Lancet, 2017)



(A) Hiérarchie traditionnelle des preuves de l'EBM.

(B) Classification par 4 niveaux de la qualité des preuves.

EMB = médecine basée sur les preuves.

GRADE = classement des recommandations évaluation, développement et évaluation.

RCT = *Randomized Controlled Trial* ou essai randomisé en double aveugle

*La qualité de l'étude diminue d'un ou deux niveaux.

**La qualité de l'étude augmente d'un ou deux niveaux.

Si l'on tente d'étendre ces réflexions aux autres professionnels de la santé et de l'éducation, cela suggère que les bonnes décisions thérapeutiques ou bonnes pratiques pédagogiques doivent se fonder non seulement sur l'ensemble de preuves disponibles (études expérimentales, observations, etc.), mais aussi sur le contexte et les caractéristiques des patients ou des élèves. En d'autres termes, ces avancées sur le EBM montrent qu'il est nécessaire de ne pas considérer les enseignants comme des « techniciens » devant appliquer mécaniquement des recommandations mais plutôt comme des « ingénieurs » et ainsi reconnaître leurs expertises et leurs savoirs dans leurs gestes professionnels. Plus généralement, les professionnels de l'enfance et de l'éducation devraient se saisir de ce débat sur les liens entre les résultats de la recherche et les recommandations cliniques et pédagogiques.

B. Les deux démarches scientifiques dans le domaine de l'éducation

Dans le domaine de l'éducation, il est classique de considérer deux types de démarches scientifiques : la première, appelée « hypothético-déductive », part de théories ou de lois aussi générales que possibles, en tire des prévisions qui sont confrontées aux données de l'expérience, la seconde, appelée « explicative » part de ce qui est, de ce que l'on observe et s'interroge sur ses déterminants en essayant de remonter la chaîne causale. Ces deux démarches sont liées et il existe entre elles des va-et-vient. Dans les deux démarches, les chercheurs ont recours à des systèmes de causalité complexes impliquant plusieurs facteurs explicatifs pouvant être mis en évidence par des analyses statistiques. Ces deux démarches sont légitimes dans le domaine de l'éducation.

La démarche « hypothético-déductive » est privilégiée dans les recherches en sciences cognitives interventionnelles. Ce type de recherche utilise en particulier la méthode dite expérimentale. Cette dernière a comme objectif principal « d'administrer la preuve » (Bernard, 1878), c'est-à-dire de montrer qu'un facteur (par exemple, une méthode d'enseignement, une technique d'apprentissage) est bien la principale cause de l'apparition d'un comportement observé (par exemple, de meilleures performances d'apprentissage), « toutes choses étant égales par ailleurs ». Pour être certain que cette relation causale est univoque, il faut organiser des « interventions » (appelées aussi entraînements, méthodes ou pédagogies) en classe, afin de contrôler au maximum ou au mieux tous les autres facteurs susceptibles d'influencer les performances observées (niveau scolaire, catégorie socioprofessionnelle, etc.). Il est bien évident qu'il s'agit d'une tâche très difficile compte tenu de la complexité du système éducatif et de son inscription dans le champ social et culturel.

Pour mesurer l'efficacité d'une « intervention ou pédagogie », on peut considérer trois niveaux de rigueur méthodologique dans le type de protocole expérimental utilisé par les chercheurs (Gentaz, 2018).

Le niveau 1 correspond aux études proposant une intervention auprès d'un groupe d'élèves (Groupe Expérimental-GE) et comparant les mesures effectuées pré- et post-intervention auprès d'un groupe d'élèves du même âge dit actif, c'est-à-dire bénéficiant également d'un programme d'intervention de nature différente à celui du premier groupe (Groupe Contrôle Actif-GCA). Avec ce type d'étude, il faut distinguer aussi le mode de sélection et d'affectation des élèves aux groupes expérimental ou contrôle. Le niveau 1A correspond aux études ayant pu effectuer une affectation aléatoire. Elles sont appelées « études randomisées en double aveugle – *Randomized Controlled Trial* – RCT ». Le niveau 1B correspond, quant à lui, aux études n'ayant pas pu le faire (par exemple, l'affectation des élèves est guidée par l'engagement volontaire de l'enseignant dans la recherche et donc des élèves de sa classe).

Le niveau 2 correspond aux études proposant une intervention à un groupe d'élèves et comparant les mesures effectuées pré- et post-intervention à un groupe d'élèves du même âge dit passif (*i.e.*, ne bénéficiant d'aucune intervention, ou placé sur liste d'attente pour bénéficier de l'intervention après l'étude ; Groupe Contrôle Passif-GCP). Si ce type d'étude permet la comparaison à un groupe de référence, il ne permet pas d'éviter d'éventuels effets psychosociaux bien connus (*e.g.*, attentes positives vis-à-vis de l'intervention, effet *placebo*) chez le groupe bénéficiant de l'intervention, effets pouvant avoir une influence sur les mesures. Comme pour le niveau 1, il faut également distinguer un niveau 2A avec une affectation aléatoire des élèves et un niveau 2B sans affectation aléatoire.

Le niveau 3 correspond aux études utilisant des protocoles « intervention seule » : l'intervention est proposée à un seul groupe d'élèves et des mesures sont proposées pré- et post-intervention pour évaluer l'impact de l'intervention mais sans comparaison à un groupe contrôle de référence. Ce type de protocole est intéressant car il permet de tester la « faisabilité » de l'intervention mais il reste exploratoire et insuffisant. En effet, il ne permet pas une interprétation fiable des résultats puisqu'aucun contrôle ne permet d'assurer que des élèves n'ayant pas bénéficié de l'intervention ne présentent pas la même évolution sur les variables mesurées. Comme pour les deux niveaux précédents, il faut aussi distinguer un niveau 3A avec une sélection aléatoire des élèves et un niveau 3B avec un choix des élèves non aléatoire.

C. Un exemple des difficultés à généraliser à grande échelle les résultats de la recherche sur l'apprentissage de la lecture en France

1. Les objectifs de l'intervention

Il existe relativement peu de recherches interventionnelles de niveau méthodologique 1 ou 2 conduites dans le système scolaire public français. Les plus nombreuses portent sur l'apprentissage de la lecture. Les résultats de la synthèse des études internationales du *National Reading Panel* (Ehri *et al.*, 2001), comme ceux des études francophones, montrent tout d'abord que, pour être efficaces, les interventions doivent être explicites, très structurées, et s'effectuer en petits groupes. Les séquences, de courte durée (20 à 30 minutes), doivent se répéter plusieurs fois dans une même semaine et ce pendant un ou deux mois. Les études conduites dans le système scolaire français montrent également que les capacités phonémiques peuvent être entraînées très tôt, en grande section de maternelle et que les interventions les plus efficaces sont celles dans lesquelles le travail oral sur les sons-phonèmes s'effectue avec le support écrit des lettres-graphèmes qui leur correspondent (Bara, Gentaz & Colé, 2004).

Encadré 1 : Graphèmes/phonèmes

Les phonèmes sont les unités de base de l'oral qui permettent de différencier, dans une langue donnée, deux mots. Le répertoire des phonèmes varie d'une langue à une autre. Par exemple, /b/ et /v/ sont deux phonèmes différents en français, permettant de différencier *bol* de *vol*, mais pas en espagnol. Les graphèmes sont les unités de base d'une écriture alphabétique qui transcrivent les phonèmes (dans *chic* : <ch>=/ʃ/, <i>=/i/, <c>=/k/). De nombreux graphèmes n'ont qu'une lettre (<a>, <k>), d'autres en ont plusieurs (<ou>, <ch>, <qu>, <eau>) et ils peuvent inclure un signe diacritique (<a>-<à>, <ou>-<où>...). Dans une écriture alphabétique, il faut mettre en relation les voyelles et les consonnes (<a>, <ou> et , <ch>) avec les phonèmes qui leur correspondent (/a/, /u/ et /b/, /ʃ/) : c'est-à-dire apprendre à *décoder*. Les recherches indiquent plus spécifiquement que les entraînements phonémiques ont des effets supérieurs sur les capacités de lecture-décodage quand ils sont associés *dans les mêmes séances* à des exercices d'exploration visuelle et tactile des lettres).

Gentaz et ses collègues (2013) ont évalué avec l'application la plus rigoureuse possible de la méthode expérimentale, à grande échelle et dans des conditions « réalistes » dictées par les contraintes des différents acteurs de l'Éducation nationale, les effets d'une intervention issue de travaux antérieurs effectués à petite échelle mais non publiés à cette époque. Cette intervention combinait des entraînements centrés sur le « décodage » et la « compréhension » sur les performances en lecture

d'enfants scolarisés en réseaux d'éducation prioritaire (REP) par rapport à une « méthode classique » (sans entraînements spécifiques). Le « décodage » est une compétence qui implique la capacité d'utiliser les correspondances graphèmes-phonèmes (et donc celle de pouvoir isoler les phonèmes), qui doit devenir progressivement automatique. La « compréhension » est une compétence qui permet d'accéder au sens d'un énoncé, qu'il soit oral ou écrit. Rappelons qu'il s'agit de la finalité même de l'apprentissage de la lecture.

2. Considérations méthodologiques et résultats

Cette intervention a été élaborée à partir d'une expérimentation préalable, intitulée « PARLER » effectuée avant 2010 auprès d'un échantillon de 107 élèves (de la grande section de maternelle au CE1) situées dans des zones d'éducation prioritaire à Grenoble et ses environs en France. Les résultats préliminaires présentés dans des conférences locales étaient très encourageants : les résultats montraient que les élèves ayant bénéficié du dispositif PARLER avaient de meilleures acquisitions en fin de CE1 dans le domaine du langage oral et de la lecture et même dans celui du raisonnement visuo-spatial non verbal ; le niveau de lecture des enfants des classes expérimentales était supérieur à la moyenne nationale. Ces résultats ont bénéficié d'une large publicité et ont évidemment ainsi intéressé les politiques français (ces résultats ne seront malheureusement publiés qu'en 2015 (Zorman *et al.*, 2015)).

Cependant, cette expérimentation présentait un certain nombre de limites : absences de répartition au hasard des élèves, d'entraînement basé sur le « décodage », de vérification des compétences dans d'autres domaines tels que les mathématiques et de contrôle des effets psychosociaux inhérents induits par une expérimentation.

En prenant en compte ces limites méthodologiques, Gentaz et ses collègues ont donc proposé une recherche interventionnelle sur une année scolaire à très grande échelle dans trois départements de l'académie de Lyon en France.

Cette recherche était singulière par son ampleur et son souci de contrôler d'éventuels biais expérimentaux comme (1) la désignation aléatoire des écoles « test » ou « témoin », (2) l'évaluation du niveau en mathématiques (pour mesurer si un enseignement massif de la lecture ne s'est pas fait au dépend des autres matières comme les mathématiques), (3) la prise en compte de tous les élèves, (4) la quasi-séparation des évaluations conjointes et complémentaires des chercheurs du CNRS et ceux de la direction de l'évaluation, de la prospective et de la performance (Depp) du ministère de l'Éducation nationale (MEN) avant et après l'intervention.

Dans les classes « test » tirées au hasard parmi un échantillon d'enseignants volontaires, ces derniers ont proposé à leurs élèves quatre entraînements centrés respectivement sur le décodage, la compréhension, la phonologie, et la fluence de lecture. La mise en œuvre de ces entraînements sur une année scolaire a nécessité une réorganisation du travail de classe. Les élèves, répartis en petits « groupes de besoin » composés de 4 à 6 enfants de niveau assez homogène, ont bénéficié de plusieurs séances structurées de 30 minutes effectuées avec leur enseignant durant chaque semaine. Pendant ces séances, les autres enfants, en petits groupes aussi, ont appris à travailler en « autonomie ». Pour dédoubler toutes les séances (celles dirigées comme celles en autonomie), chaque classe a bénéficié d'un adulte supplémentaire. Dans les classes témoins (tirées au hasard parmi un panel de volontaires), les enseignants ont eu comme consigne de ne pas modifier leurs pratiques ordinaires de classe.

Pour évaluer les progrès en lecture, les performances de 1 252 élèves de CP des classes « test » ont été mesurées avant et après les séances d'entraînement et ont été comparées à celles de 2 398 élèves des classes « témoin ». En parallèle, la Depp du MEN a conduit sa propre évaluation (lecture et mathématiques) avec sa procédure habituelle (livrets d'évaluation distribués aux enseignants) dans les mêmes classes ainsi que dans d'autres classes similaires issues de toute la France en testant au final 2 375 élèves.

Les analyses quantitatives des résultats aux évaluations conduites par les chercheurs du CNRS et ceux de la Depp montrent de manière similaire que le programme n'a pas permis de faire davantage progresser en lecture les enfants du groupe « test » (académie Lyon) que ceux des groupes « témoin » (académie de Lyon et niveau national). En revanche, les évaluations subjectives réalisées par les différents inspecteurs de l'Éducation nationale (IEN) auprès des enseignants impliqués dans le programme et exposées trimestriellement au cours des réunions du comité de pilotage réunissant tous les acteurs du projet (recteur, inspecteur d'académie, etc.) étaient très positives, à ce à différents niveaux : comportements individuels et collectifs des enfants, ambiance de classe, travail d'équipe.

3. Limites

Cette recherche interventionnelle présente aussi plusieurs limites (Gentaz *et al.*, 2013). Compte tenu du fonctionnement du MEN dans ses processus de décision et de financements, les chercheurs n'ont pu disposer que de peu de temps pour préparer le projet, puis pour recruter et former les enseignants volontaires avant le début de l'année de l'expérimentation. Plusieurs mois sont nécessaires pour former les enseignants aux entraînements et nouvelles organisations de classe à mettre en œuvre. Des études (Blachman, Ball, Black & Tangel, 1994) montrent des différences entre les recherches réalisées dans la situation ordinaire de la classe et celles effectuées par des chercheurs, le point capital étant que les enseignants impliqués doivent être bien formés (Bara *et al.*, 2007). Cette limite temporelle dans sa mise en place et sa conduite a aussi eu pour conséquence de limiter fortement la possibilité de proposer une intégration des contenus des quatre entraînements et de les coordonner avec les autres outils utilisés par les enseignants (manuels de lecture, fichiers, etc.). Elle a ainsi empêché, comme noté dans le rapport de l'inspection générale du MEN en France (inspection générale, 2011), de corriger certaines des imperfections des outils utilisés pour cette expérimentation, outils qui avaient été développés avant elle (par exemple, celui dédié à l'entraînement phonologique). De plus, on ne peut pas exclure qu'un manque de synergie entre les entraînements spécifiques et les différents et nombreux manuels de lecture utilisés en classe puisse expliquer, en partie, les moins bonnes performances des enfants du groupe « test ».

La seconde limite est liée aux conséquences de l'ampleur de l'étude sur le dispositif de formation et de suivi des enseignants : du fait de leur grand nombre et de leur dispersion géographique sur trois départements (contraintes imposées par le rectorat), les chercheurs ont eu des difficultés à proposer une formation approfondie régulière et un suivi personnalisé pour chacun des enseignants impliqués. Au total, chaque enseignant n'a pu recevoir qu'environ 30 heures de formation réparties sur l'année et ce seulement en dehors de la classe, et non *in situ*.

La troisième limite est qu'il n'a été possible d'intervenir que durant la première année du primaire pour des raisons financières politiques. Or, les études montrent l'importance de la préparation à la

lecture dès la maternelle et la nécessité d'avoir des entraînements se déroulant sur 2 ou 3 années pour observer des effets bénéfiques significatifs.

4. Conclusions

Ce type de recherche pose de réelles questions théoriques et pratiques concernant le transfert et la généralisation des résultats obtenus avec des petits groupes de quelques dizaines d'enfants à des grands groupes de centaines d'enfants. Cette recherche illustre parfaitement le débat ancien sur l'utilisation directe des résultats de la recherche. C'est ce que signale cette étude, qui montre que, si les résultats sur l'apprentissage de la lecture issus des sciences cognitives sont solides, et si les principes pédagogiques qui en découlent sont aujourd'hui bien connus, leur mise en application dans les classes à grande échelle demande encore un effort très important de recherche devant mobiliser conjointement les acteurs du monde de la recherche et ceux du monde éducatif (Deheane, Dehaene-Lambert, Gentaz, Huron & Sprenger-Charolles, 2011).

En 2013, l'institut des sciences de l'éducation du département états-uniens d'éducation a publié une note faisant une synthèse des 90 recherches randomisées qu'il avait engagées depuis 2002 pour évaluer l'efficacité de diverses interventions ayant implanté à grande échelle des programmes, pratiques ou stratégies éducatives en comparaison avec les pratiques habituelles (IES, 2013). L'analyse finale révèle que 77 des 90 recherches randomisées se sont révélées ne comportant aucun défaut méthodologique majeur. Sur ces 77 études, 7 interventions (9 %) produisent des effets positifs significatifs et 70 ne produisent que des effets faibles ou ne produisent aucun effet positif. Les auteurs précisent que ces résultats ne sont pas propres aux recherches en éducation.

En conclusion, pour tenter de mieux comprendre ces résultats décevants obtenus à grande échelle, il existe des recherches récentes portant sur les processus d'implémentations dans le domaine de la santé qui pourraient être éclairantes aussi pour le domaine de l'éducation.

II. Les apports des sciences de l'implémentation des interventions

A. Facteurs influençant le processus d'implémentation dans le domaine de la santé

La revue critique réalisée par Durlak et DuPre (2008) vise à évaluer l'impact de l'implémentation sur les effets générés par des programmes de prévention et de promotion de la santé pour les enfants et les adolescents. Les résultats indiquent clairement qu'une implémentation efficace d'un programme est associée à de meilleurs résultats.

Les données de 483 études évaluées dans cinq méta-analyses indiquent que l'amplitude de la moyenne des tailles d'effet de programmes interventionnels est au moins deux à trois fois plus élevée lorsque les programmes sont soigneusement implémentés et ne posent pas de graves problèmes d'implémentation que lorsque ces conditions ne sont pas respectées.

Les données de 59 études quantitatives supplémentaires confirment que des niveaux plus élevés d'implémentation sont souvent associés à de meilleurs résultats, en particulier lorsque la fidélité ou le dosage (*cf.* définitions données plus loin) sont évalués.

Il existe donc un certain nombre de preuves convaincantes qui confirment l'impact capital de l'implémentation d'un programme sur les résultats obtenus suite à l'implémentation d'un programme. Les auteurs ajoutent qu'en l'absence de données sur l'implémentation, la recherche ne peut documenter avec précision comment les résultats devraient être interprétés.

L'évaluation de l'implémentation d'un programme semble par conséquent être une nécessité absolue. Les auteurs relèvent toutefois qu'il n'est pas réaliste d'attendre que les praticiens implémentent de manière parfaitement fidèle à 100 % l'ensemble d'un programme. D'ailleurs, aucune étude n'a relevé un niveau d'implémentation correspondant à 100 %. Avec des niveaux d'implémentation de 60 % à 80 %, des résultats positifs sont souvent obtenus. Cette marge d'adaptation fait donc partie intégrante de l'intégration d'outils innovants.

En se basant sur 81 études comprenant des données quantitatives ou qualitatives relatives aux facteurs influençant le processus d'implémentation, les auteurs ont pu mettre en évidence 23 facteurs regroupés dans 5 grandes catégories qui peuvent expliquer ces variations dans la mise en œuvre d'un dispositif :

- **Catégorie 1 : Facteurs « Macro »**
Ces facteurs sont généraux de type « communautaire », tels que le système de recherche sur la prévention, les politiques, le financement et les politiques sociales (*social policy*).
- **Catégorie 2 : Praticien.** Cette catégorie comprend les caractéristiques relatives au « praticien » du programme, à savoir ses perceptions relatives aux besoins et bénéfices d'une innovation, le sentiment d'efficacité à pouvoir mettre en œuvre ce qui est attendu et la possession des compétences nécessaires à l'implémentation.
- **Catégorie 3 : Caractéristiques de l'innovation**
Cette catégorie comprend deux grands axes : la compatibilité et l'adaptabilité. La compatibilité permet de prendre en considération le degré de correspondance entre l'intervention proposée et

les valeurs, missions, priorités d'une institution. L'adaptabilité prend en compte les modifications qui peuvent être réalisées sur un programme en fonction notamment des valeurs, des normes culturelles du praticien.

- **Catégorie 4 : Capacité organisationnelle**

Cette catégorie met en lumière l'importance notamment du « climat de travail », de l'ouverture à de nouvelles approches ; de la manière de gérer le groupe (*e.g.*, en déterminant les priorités, en établissant un consensus, etc.) ; d'une personne ressource respectée par l'équipe qui « défend » l'outil innovant, trouve des solutions ; et de l'importance de l'implication des praticiens dans la démarche d'implémentation. L'implémentation d'une intervention devrait pouvoir miser sur des prises de décisions partagées entre les différents partenaires (par ex., entre les chercheurs et les praticiens). Selon Durlak et DuPre (2008), il importe de pouvoir donner davantage de responsabilités (*Empowerment*) aux différents membres de la communauté impliqués dans la mise en œuvre d'une nouvelle intervention. « *Idéalement, ce processus collaboratif est caractérisé par une confiance mutuelle et une communication ouverte, le partage de responsabilités pour réaliser des tâches importantes, et les efforts pour arriver à un consensus lorsque des désaccords ou des impasses surviennent* » (notre traduction, p. 338). Les méthodes collaboratives s'avèrent ainsi plus efficaces pour implémenter et maintenir une nouvelle intervention. En effet, les données de plusieurs études qualitatives et quantitatives portant sur des facteurs comme l'adaptabilité des programmes et la prise de décision partagée suggèrent qu'une meilleure implémentation a lieu lorsque les praticiens peuvent apporter certains ajustements aux programmes.

Dusenbury, Brannigan, Falco et Hansen (2003) suggèrent d'une part que les praticiens pourraient améliorer l'outil en lui-même. Les adaptations réalisées par les praticiens seraient plus sensibles et adaptées au contexte particulier de leur classe. En effet, des protocoles extrêmement détaillés peuvent être faisables dans des recherches très contrôlées en laboratoire, mais peuvent poser un certain nombre de difficultés lors de leur implémentation sur le terrain dans des conditions de travail ordinaires. D'autre part, il est également possible que les praticiens qui adaptent une intervention soient plus créatifs et motivés de manière générale et s'avèrent être au quotidien de meilleurs praticiens. Dans tous les cas, les retours et commentaires des praticiens par rapport au contenu des outils à mettre en œuvre constituent une source d'information précieuse à prendre en considération afin d'optimiser l'implémentation d'un programme.

- **Catégorie 5 : Formation et l'assistance technique**

Cette catégorie comprend une formation qui englobe à la fois le soutien au développement d'aptitudes spécifiques à l'intervention, mais également le soutien à la motivation, la prise en compte des attentes et du sentiment d'auto-efficacité. Les auteurs relèvent l'importance des « formes actives » d'apprentissage. Les formations qui s'avèrent efficaces comprennent, des jeux de rôles, du modelage et des retours d'information sur ce qui a été mis en œuvre. L'assistance technique comprend « *les ressources offertes aux praticiens une fois l'intervention commencée* » (Dusenbury *et al.*, 2003, notre traduction, p. 339). Ces ressources comprennent par exemple le soutien à la motivation des personnes qui implémentent un programme, la résolution de problèmes pouvant survenir lors de l'implémentation, l'amélioration des compétences des praticiens, le soutien émotionnel etc.

B. Définition et composantes liées à l'implémentation

En tant que discipline, l'apparition des sciences de l'implémentation est liée au fait que la recherche produisait des preuves de l'efficacité de certaines pratiques, mais que les praticiens ne les appliquaient pas dans leurs propres programmes. Cette très jeune discipline s'est formalisée au cours des deux dernières décennies avec la publication de la revue « *Implementation Science* », en 2006, la formation de la « *Global Implementation Society*³ » et de la « *Society for Implementation Research Collaboration*⁴ » qui visent à promouvoir et à faciliter la communication et la collaboration entre les pratiques, sciences et politiques de l'implémentation (Odom, Hall & Suhrheinrich, 2020). Deux pionniers de la science de l'implémentation (SI), Eccles et Mittman (2006) ont défini la SI comme « *l'étude scientifique des méthodes visant à promouvoir l'adoption systématique des résultats de la recherche et d'autres pratiques fondées sur des données probantes dans la pratique courante, et donc à améliorer la qualité et l'efficacité des services de santé.* ». Plus spécifiquement, le terme « implémentation » fait référence « *au processus de mise en place d'une pratique ou d'un programme dans le fonctionnement d'une organisation, telle qu'une école, et peut être considérée comme l'ensemble des activités conçues pour y parvenir.* » (Forman et al., 2013, notre traduction, p. 78)

Actuellement, il est possible de distinguer un certain nombre de composantes liées à la notion d'implémentation que les chercheurs peuvent évaluer dans le cadre de la mise en œuvre d'un outil innovant.

Proctor et ses collègues (2011) proposent une taxonomie de « résultats liés à l'implémentation ». Ils définissent les « résultats de l'implémentation » de la manière suivante : « *les effets d'actions délibérées et planifiées pour implémenter de nouveaux traitements, de nouvelles pratiques et de nouveaux services.* » (notre traduction, p. 65). Nous les décrivons succinctement ci-dessous :

- **L'acceptabilité** renvoie à la manière dont les praticiens/intervenants perçoivent l'implémentation d'un traitement, d'un service, d'une pratique ou d'une innovation spécifique, si cette dernière est perçue comme étant agréable, acceptable ou satisfaisante.
- **L'adoption** est définie comme « l'intention, la décision initiale ou l'action d'essayer ou d'utiliser une innovation ou une pratique fondée sur des preuves » (notre traduction, p. 69).
- **L'adéquation** réfère notamment à la compatibilité de l'intervention ou de la pratique fondée sur des données probantes pour un contexte de pratique précis, un intervenant ou un consommateur donné. Ce type de résultat permet d'estimer si le dispositif innovant est perçu comme étant adéquat pour traiter une question ou un problème particulier.
- **Le coût** (coût « incrémental » / supplémentaire ou coût d'implémentation) est défini comme le coût spécifique généré par l'implémentation.

³ <https://globalimplementation.org/>

⁴ <https://societyforimplementationresearchcollaboration.org/>

- **La faisabilité** est considérée comme une mesure qui va permettre d'estimer si un nouveau traitement ou une innovation (*e.g.*, outil didactique) peut être utilisé ou réalisé avec succès dans un contexte donné.
- **La fidélité** renvoie au degré auquel une intervention a été implémentée telle qu'elle était prescrite dans le protocole original ou telle qu'elle était développée par les concepteurs du programme. De nombreuses définitions de la fidélité d'implémentation existent. C'est pour cette raison qu'il importe d'être très précis dans les problèmes spécifiques liés à la fidélité qu'une recherche pourrait rencontrer (Dusenbury *et al.*, 2003)
- **La pénétration** réfère à l'intégration d'une pratique, d'un dispositif au sein d'un milieu de services, comme dans le contexte scolaire, et de ses sous-systèmes. Elle permet de mesurer le « degré d'institutionnalisation d'un traitement donné » (nombre d'intervenants qui mettent en œuvre le programme divisé par le nombre d'intervenants qui devraient le mettre en œuvre).
- Enfin, **la durabilité** nous donne des informations utiles relatives au maintien ou à l'institutionnalisation d'un programme nouvellement implémenté dans le cadre des activités ordinaires d'un établissement de services.

Parmi l'ensemble de ces différents paramètres permettant d'évaluer l'impact de l'implémentation d'une intervention, la fidélité d'implémentation est plus souvent mesurée contrairement aux autres indicateurs. Les cinq dimensions constitutives de la fidélité d'implémentation seront par conséquent décrites dans la section suivante.

C. La mesure de la fidélité d'implémentation

Comme évoqué précédemment, assurer l'implémentation d'un programme ou d'un curriculum constitue une idée centrale dans la perspective de soutenir la transition d'un programme planifié (prescrit, développé par les chercheurs) au programme réellement mis en œuvre dans un contexte particulier. Cependant, lorsqu'un outil didactique n'exerce pas l'effet attendu sur les apprentissages des élèves, ce résultat est souvent (à tort) interprété comme si le programme ou le curriculum ne « fonctionnait pas », entraînant son abandon (Fisher, Smith, Finney & Pinder, 2014). Ces auteurs suggèrent alors qu'il serait prématuré de conclure que les effets du programme sont invalides, sans données permettant d'évaluer précisément la fidélité dans la mise en œuvre d'une intervention. La fidélité d'implémentation est abordée dans de nombreux domaines (éducation, santé, psychologie) et vise à examiner le degré auquel un programme est exécuté tel qu'il a été planifié (prescrit).

Il existe cinq composantes clés identifiées et définies par Dane et Schneider (1998) sur lesquelles prendre appui pour collecter des données sur la fidélité d'implémentation. Ces composantes ont été par la suite reprises par Fisher *et al.* (2014) et par Gerstner et Finney (2013) qui présentent une approche pratique pour évaluer la fidélité d'implémentation.

Nous nous fondons principalement sur l'article de Gerstner et Finney (2013) pour les expliciter brièvement :

- **Spécificité du programme.** Ce critère permet de répondre à la question suivante : « Quelles sont les composantes et les caractéristiques spécifiques du programme ? ». Ces composantes et caractéristiques constitutives de l'outil devraient théoriquement permettre aux enfants d'atteindre les objectifs d'apprentissage visés par le programme. Même si ce critère n'est pas directement « évalué », il constitue l'aspect le plus fondamental de l'évaluation de la fidélité, car c'est sur la base de ce critère que les critères « d'adhésion » et de « qualité » vont être évalués. Il devrait par conséquent être décrit de manière détaillée.
- **Adhésion.** Cette composante renvoie à ce type de question : « Est-ce que les caractéristiques qui font la particularité du programme ont été implémentées telles que prévu ? ». Par le biais de ce critère, le chercheur évalue si les spécificités du programme ont été implémentées et si par conséquent les élèves ont eu l'opportunité d'acquérir les habiletés et les connaissances leur permettant d'atteindre le ou les objectifs fixés. Il est possible d'évaluer ce critère en demandant aux personnes implémentant le programme d'indiquer si ce qu'elles ont mis en œuvre correspond aux caractéristiques spécifiques du programme.
- **Qualité.** Ce critère traduit dans une certaine mesure l'efficacité à laquelle l'enseignant/intervenant a conduit le programme. « Est-ce que la qualité d'implémentation du programme est élevée ? ». En d'autres termes, « est-ce que l'enseignement prodigué est de grande qualité ? ». Selon les auteurs, les intervenants peuvent tout à fait mettre en œuvre les différentes étapes de la planification (haute adhésion), mais l'exécution des différentes étapes planifiées peut être médiocre. Par exemple, certaines consignes, certaines relances ou questions sont formulées trop lentement, trop rapidement ou pas au moment opportun (inattention des élèves).
- **Exposition ou dosage.** Cette quatrième composante évalue dans quelle mesure tous les élèves participant à un programme bénéficient de l'ensemble de l'intervention. Cette composante est étroitement liée au premier critère, dans la mesure où la spécificité du programme rend également visible la durée de chaque intervention. Les auteurs indiquent qu'en plus d'évaluer la durée du programme, ce critère évalue si chaque enfant a pu bénéficier de l'entièreté de l'intervention. Pour Dane et Schneider (1998), ce critère inclut également le nombre de séances, la longueur de chaque séance et la fréquence à laquelle les techniques utilisées sont mises en œuvre.
- **Réceptivité.** Cette dernière composante renvoie à la réceptivité des personnes, et ici en l'occurrence des enfants qui sont exposés au programme. Ce critère réfère par conséquent au degré d'intérêt, d'engagement, de compréhension des élèves lors des séances d'enseignement/apprentissage.

Ces différentes composantes, qui peuvent être envisagées comme des critères, combinées aux effets d'un programme sur les variables mesurées dans divers domaines (émotionnels, sociaux, scolaires, etc.) devraient faire systématiquement l'objet d'une évaluation lors de l'implémentation d'un outil innovant et ce afin de fournir d'intéressantes hypothèses explicatives quant à l'efficacité d'un programme par rapport aux objectifs poursuivis (faible vs. haute-fidélité combinée aux effets favorables vs défavorables du programme, voir les quatre propositions de scénarios de Gerstner et Finney, 2013).

D. Exemple d'une démarche de conception d'outils didactiques en France

Selon les auteurs (Cèbe & Goigoux, 2018) de la démarche de conception d'outils didactiques que nous allons vous présenter, dans la perspective d'améliorer l'enseignement, faire usage de nouveaux outils didactiques ou de programmes d'intervention comprenant des dispositifs innovants constitue un formidable moyen de favoriser le développement professionnel des enseignants. Toutefois, afin d'optimiser la conception et l'implémentation d'outils en classe, il s'agit d'associer dès le départ les enseignants dans la conception d'outils didactiques innovants.

Cèbe et Goigoux (2018) rejoignent ainsi Durlak et DuPre (2008) ou Buysse, Sparkman et Wesley, (2003) en insistant sur le fait que le partenariat entre chercheurs et enseignants sur le terrain favoriserait la création de nouveaux outils « cohérents avec les résultats de la recherche (le 'souhaitable'), mais compatibles avec les pratiques habituelles des enseignants (le 'raisonnable') » (p. 80).

Cèbe et Goigoux (2018) proposent trois caractéristiques qui devraient permettre d'optimiser la formation des praticiens par les outils didactiques (dont les programmes d'intervention font partie) :

1. Leur conception doit pouvoir prendre en compte les questionnements et préoccupations des praticiens sur le terrain et répondre aux besoins d'enseignement et d'apprentissage des élèves qui en bénéficieraient le plus.
2. Leur implémentation en classe ne devrait pas exagérément bouleverser les conditions de travail ordinaire des enseignants.
3. Leurs fondements théoriques et empiriques devraient être explicités afin que les enseignants puissent saisir les cadres de référence qui les sous-tendent et justifient de la mise en œuvre des outils proposés.

Sur la base de ces différents constats, les auteurs proposent une « démarche de conception d'outils didactiques » tout à fait intéressante que nous allons décrire ci-dessous.

La démarche de conception que les auteurs nomment « conception continuée dans l'usage » peut être décrite comme un processus qui se veut itératif et collaboratif comprenant trois étapes que nous allons décrire succinctement :

- **Étape 1 : la création d'un premier prototype**

Cette étape comprend l'analyse des données scientifiques disponibles et l'établissement d'un premier compromis entre les résultats de la recherche et les pratiques quotidiennes des enseignants. Ce sont les chercheurs qui initient cette démarche en prenant appui sur trois grands types de connaissances relatives, 1) aux processus d'apprentissage des élèves dans un domaine particulier, quels sont les éléments de compétences à développer ? 2) aux techniques d'enseignement qui s'avèrent efficaces au regard de la recherche dans le domaine concerné, quelle démarche pédagogique retenir ? 3) aux pratiques d'enseignement ordinaires.

- **Étape 2 : l'implémentation de deux prototypes dans les classes, « la co-conception »**

Cette étape va permettre de tester le premier prototype afin d'élaborer le second prototype, par suite, de tester le second prototype pour créer le dernier prototype. Le nombre d'enseignants augmente au fur et à mesure de l'implémentation des prototypes (un groupe d'enseignants pour

le premier prototype, puis deux groupes d'enseignants incluant le premier groupe plus une nouvelle équipe d'enseignants). Les modifications sont également négociées avec les enseignants.

- **Étape 3 : « l'évaluation »**

Au terme des deux premières étapes, les effets du troisième prototype sur les apprentissages des élèves sont évalués à grande échelle. Les progrès des élèves sont alors comparés dans des conditions contrôles et expérimentales.

Une quatrième étape nécessite toutefois d'être ajoutée. Nous l'intitulerons de la manière suivante : « Éclairer les politiques éducatives ». Cette phase incontournable à la pérennité d'un programme renvoie à l'intégration des outils innovants ayant attesté de leur « efficacité » aux réformes dans les politiques éducatives d'un district, d'un ou plusieurs canton(s), voire d'un pays. Ces réformes peuvent alors engendrer un remaniement du Plan d'Etudes de la formation de l'élève. En effet, obtenir des résultats probants est nécessaire, mais non suffisant si le programme ne fait pas l'objet d'une plus large implémentation dans les écoles, grâce à une diffusion à grande échelle « validée » (« *scale-up process* », (Klingner, Boardman & McMaster, 2013) et fortement encouragée par une instance hiérarchique supérieure comme par exemple en Suisse la CIIP (Conférence intercantonale de l'instruction publique de la Suisse romande et du Tessin). Cette dissémination nécessite notamment d'être prise en charge par des formateurs et des animateurs pédagogiques formés à l'implémentation de l'outil innovant qui peut alors devenir un nouveau moyen d'enseignement sur lequel l'enseignant pourra prendre appui en classe pour soutenir les apprentissages de ses élèves. Si cette transition entre l'étape 3 et 4 n'est pas anticipée ou n'a pas lieu, il y a alors de fortes chances que l'usage d'un programme ne demeure circonscrit localement, voire disparaisse. Comme le relève Klingner *et al.* (2013), après le départ des chercheurs et le retrait des financements liés à la recherche, la mise en œuvre d'un programme dans les classes a tendance à faiblir.

E. Exemple d'une recherche interventionnelle visant à développer les compétences socio-émotionnelles chez des enfants de 5-6 ans : processus d'implémentation d'un programme en contexte scolaire dans le Canton du Valais en Suisse.

Dans cette section, nous allons vous présenter une recherche comprenant trois expérimentations en contexte scolaire dont l'objectif principal était d'examiner les effets de l'implémentation d'un programme basé sur le jeu symbolique ou de faire semblant (un jeu dans lequel l'enfant crée une situation imaginaire comprenant la sélection et l'interprétation de rôles au travers desquels il s'applique des règles de comportement correspondant aux rôles qu'il endosse (Vygotsky, 1933/2016)) dans le but de promouvoir certaines compétences socio-émotionnelles.

Le programme expérimenté dans les classes comprend des sessions de jeux de faire semblant associées à des phases d'enseignement/apprentissage plus systématisées avec le collectif d'apprentissage autour de compétences socio-émotionnelles spécifiquement délimitées. Tout comme dans la démarche de conception d'outils didactiques proposée par Cèbe et Goigoux (2018) (sur laquelle nous nous appuyons ici pour décrire le processus d'implémentation de nos différentes séances) l'ensemble du processus d'implémentation du programme comprend trois grandes phases correspondant à trois études différentes que nous allons présenter succinctement.

- **Étape 1 : étude de faisabilité du programme (Richard, Gay, Clerc-Georgy & Gentaz, 2019)**

La première étape a consisté à dresser un état des lieux des manières privilégiées d'apprendre dans les premiers degrés de la scolarité et par conséquent des modalités de travail à mettre en œuvre par l'enseignant afin de soutenir les progrès des élèves. Le jeu et tout particulièrement le jeu de faire semblant (plus spécifiquement le jeu socio-dramatique) s'est très rapidement imposé comme un moyen privilégié pour contribuer au développement des compétences cognitives (par exemple, le contrôle inhibiteur, White, Thibodeau-Nielsen, Palermo & Mikulski, 2021) et socio-émotionnelles des élèves de 5-6 ans (pour une synthèse d'études quantitatives sur les liens entre compétences socio-émotionnelles et jeu de faire semblant, cf. Richard & Gentaz, 2020). Dans un second temps, nous nous sommes ensuite appuyés sur des programmes scolaires universels ciblant à la fois l'apprentissage social et émotionnel, mais soutenant également le développement du jeu de faire semblant. À notre connaissance, seuls le programme populaire *Tools of the Mind* développé par (Bodrova & Leong, 2001) et le programme développé dans la thèse de (Landry, 2014) réalisent un accompagnement spécifique en contexte scolaire autour du jeu de faire semblant dans le but de stimuler le développement du jeu en lui-même, mais également celui de différentes compétences (notamment l'auto-régulation pour le programme *Tools of the Mind*). Nos séances ont par conséquent été en partie adaptées de Landry (2014) qui s'appuie sur Bodrova et Leong (2012) et le programme *Second Step* (Committee-For-Children, 2012). Au cours de l'étape 1, la chercheuse principale qui est enseignante et psychologue a pu tester le premier prototype du programme comprenant onze séances hebdomadaires d'environ 60 minutes dans une classe (condition expérimentale) pendant qu'un autre groupe d'élèves (condition contrôle) poursuivait le cursus scolaire habituel comprenant également du jeu de faire semblant, mais ne bénéficiant pas spécifiquement de l'intervention en classe.

Cette étude exploratoire a permis à la chercheuse d'évaluer notamment la faisabilité du programme, mais également d'examiner les effets de l'outil didactique sur le développement de la compréhension, de la régulation des émotions et du comportement prosocial chez 19 enfants âgés de 5 à 6 ans (Richard *et al.*, 2019). Les résultats ont alors montré pour le groupe expérimental, comparativement au groupe contrôle, une amélioration de la capacité à reconnaître les émotions de colère, de dégoût, les émotions présentées avec un contexte visuel et de la capacité à associer un terme émotionnel à une expression faciale émotionnelle, et une diminution de la fréquence d'utilisation des stratégies de régulation dysfonctionnelles. Suite à cette première expérimentation, un grand nombre d'éléments ont été conservés (activités proposées, accessoires utilisés, consignes, relances formulées etc.), d'autres aspects ont fait l'objet de changements (par ex., durée des activités au sein de certaines séances, fractionnement des deux premières séances en deux fois 30 minutes, suppression/ajout d'activités ou d'histoires lues aux enfants, création d'illustrations, adaptation de certaines consignes).

- **Étape 2 : étude de la conduite du programme par les enseignants (Richard *et al.*, 2021)**

Au cours de la seconde phase, le programme a été introduit dans cinq classes de deuxième année (5-6 ans) et il a été conduit par les enseignantes (Richard *et al.*, 2021). Dix classes ont été recrutées, soit cinq classes composant le groupe expérimental (n = 39) et cinq classes composant le groupe témoin (n = 40). Dans le canton du Valais en Suisse, les enfants de 5-6 ans venaient également l'après-midi à l'école sans les enfants de 4 ans dans les classes testées. Chaque classe comprend en moyenne 10 élèves de 4 ans et 10 élèves de 5-6 ans et les effets du programme uniquement ont été évalués seulement sur le groupe des 5-6 ans. Les résultats pour cette étude ont permis de reproduire en partie

les résultats obtenus dans l'étude 1, à savoir une amélioration significativement plus importante pour le groupe expérimental du vocabulaire émotionnel et de la capacité à reconnaître l'émotion de colère. Dans cette étude, comparativement à l'étude 1, les enfants ayant bénéficié du programme étaient également meilleurs dans la reconnaissance des émotions de manière globale et sans contexte visuel. Malheureusement, ni l'amélioration de la reconnaissance des émotions en contexte visuel, ni une diminution significative de la fréquence d'utilisation des stratégies de régulation dysfonctionnelles en faveur du groupe expérimental n'ont pu être reproduits.

Tout au long du processus d'implémentation du programme (avant, pendant et après implémentation), les enseignantes du groupe expérimental ont bénéficié de 20 heures de formation avec la chercheuse principale. Le contenu de ces temps de formation portait à la fois sur les fondements théoriques et empiriques qui justifient l'élaboration du contenu des différentes séances, mais également sur le processus d'implémentation du programme. En effet, pendant et après la réalisation du programme, des séances de « *debriefing* » étaient planifiées afin non seulement de soutenir la motivation des enseignantes, mais également dans le but de pouvoir prendre en considération leur point de vue par rapport aux contenus du programme et à ses effets sur les progrès des élèves. Les discussions entre les enseignantes et la chercheuse portaient alors sur leurs observations quant à l'influence des séances sur les apprentissages des élèves, sur leurs questionnements, leurs doutes. Ces rencontres ont également permis de négocier certaines modifications à apporter au programme, comme le fait d'ajouter un temps supplémentaire pour permettre aux enfants d'explorer les accessoires de jeu ou de rendre une activité supplémentaire « facultative » (dépendante du niveau d'appréciation de l'enseignant quant à l'attention des élèves au cours de la séance). C'est également suite aux retours des enseignantes que l'ensemble des illustrations utilisées ensuite dans l'étude 3 a pu être finalisé. Enfin, c'est au cours de ces moments d'échanges avec les enseignantes que la chercheuse a pris conscience du fait qu'un document proposant des pistes d'intervention spécifiques (en plus des indications présentes dans les séances) pour soutenir le développement du jeu de faire semblant manquait. Ce document (livret pratique) a été élaboré ultérieurement par la chercheuse principale et il a pu être présenté à l'équipe d'enseignants participant à l'étude 3.

• **Étape 3 : étude à grande échelle (Richard, Clerc-Georgy & Gentaz, soumis)**

Enfin, lors de la troisième étape, la dernière version du programme a été introduite durant l'année scolaire 2020-2021 à plus grande échelle ($n = 180$) dans 12 classes comprenant des enfants âgés de 5 à 6 ans (Richard, Clerc-Georgy & Gentaz, en préparation). Douze autres classes faisaient partie du groupe témoin ne bénéficiant pas de l'intervention. Les enseignants de l'équipe implémentant le programme ont de nouveau bénéficié d'environ 20 heures de formation. Toutefois, au cours de cette étude, des mesures visant à évaluer certains paramètres de l'implémentation du programme ont été utilisées. Pour ce faire, les enseignants du groupe expérimental ont rempli une grille hebdomadaire comprenant des questions sur la fidélité d'implémentation, l'exposition des enfants au programme ainsi que la réceptivité des élèves aux contenus ayant fait l'objet d'un enseignement/apprentissage. Lors de l'implémentation des 11 séances, la chercheuse principale ainsi que deux auxiliaires de recherche sont allées en classe observer 14 séances afin d'évaluer la fidélité d'implémentation ainsi que la réceptivité des élèves pour le programme. Le degré de concordance entre l'évaluation, des enseignantes et des observateurs externes a pu être estimé. Au terme de l'expérimentation en classe les enseignants ont rempli un questionnaire permettant aux chercheurs de compléter les mesures

relatives à l'implémentation de l'outil didactique (par ex des questions sur la pérennité du programme ou sa faisabilité ont été ajoutées). Les résultats concernant les progrès réalisés par les élèves sont en cours d'analyse.

Conclusion, limites et perspectives

L'administration des preuves scientifiques en éducation « *Evidence-Based Practice* » représente actuellement une étape incontournable à la mise en œuvre d'interventions dans les écoles. Cependant, la difficulté à pouvoir généraliser à grande échelle les résultats d'une recherche avec une méthodologie expérimentale rigoureuse constitue une limite centrale à ce type de démarche. Des efforts importants de recherche nécessitent encore d'être développés afin d'examiner, lors de la mise en œuvre de protocoles expérimentaux, les conditions qui favorisent l'efficacité des interventions implémentées par les enseignants à grande échelle. Dans le cadre de ce rapport, nous avons justement tenté de décrire les apports essentiels des sciences de l'implémentation des interventions afin de mieux circonscrire les facteurs multiples qui influencent la mise en place d'un programme dans les classes.

Pour conclure, un certain nombre de constats centraux explicités ici sous la forme de points de vigilance nécessitent d'être mis en lumière au terme de cette contribution, car il constitue des pistes de recherche qu'il s'agit de renforcer dans le domaine des sciences de l'implémentation en contexte scolaire et en particulier afin de soutenir ce délicat passage à grande échelle.

A. La démarche collaborative

Le premier constat évoqué déjà à plusieurs reprises concerne la nécessité de pouvoir en tant que chercheur/concepteur d'un programme demeurer ouvert aux propositions, préoccupations qui émanent du terrain. Dans ce sens, nous nous rallions à la perspective défendue par Buysse *et al.* (2003) quant à l'importance de stimuler les échanges entre chercheurs et praticiens autour des manières de connecter ce que l'on sait « *What we know?* » par le biais de la recherche avec ce qui est fait « *What we do?* » sur le terrain (notion de communauté de pratiques). Au travers de nos trois études portant sur l'impact d'un programme basé sur le jeu de faire semblant, nous avons réellement pu éprouver l'importance de cette démarche collaborative afin d'effectuer certains changements plus adaptés au contexte particulier dans lequel se déploie un outil didactique, mais également pour maintenir l'implication des enseignants dans la mise en œuvre complète de l'intervention.

B. L'appropriation d'un outil innovant

Le second élément qui nous semble essentiel à mettre en évidence concerne la « spécificité » des programmes ainsi que leur appropriation par l'enseignant. Selon Gerstner et Finney (2013), cet aspect de la fidélité d'implémentation, à savoir la « spécificité » de l'outil, nécessite de la part du concepteur de l'intervention de détailler les composantes clés du programme. Durlak et DuPre (2008) relèvent également que les interventions qui conduiraient à un plus haut niveau d'implémentation sont accompagnées de plans de cours détaillés. Toutefois, il importe certes de pouvoir fournir des instructions précises et complètes, mais il s'agit également de pouvoir trouver un dosage équilibré entre excès et manque d'informations. En effet, pour certains enseignants participant à nos études (par ex., Richard *et al.*, en préparation), la planification trop détaillée de certaines séances surchargeait la mise en œuvre de l'intervention (gestion du nouveau matériel, des élèves, des questions, des relances tout en observant les élèves) et ne leur permettait pas d'appréhender la séance de manière complètement « naturelle » et spontanée (besoin de lire les consignes ou d'avoir la planification en permanence sous les yeux). Nous formulons l'hypothèse que les enseignants ont besoin d'un temps

d'appropriation du nouveau dispositif et que potentiellement les enseignants intégrant un nouvel outil en classe auront une meilleure maîtrise de ce dernier les années ultérieures. D'ailleurs, Nesbitt et Farran (2021) nous rejoignent sur ce point en précisant que généralement les études de type randomisées contrôlées évaluent les effets d'interventions dans des classes où l'enseignant met en œuvre de nouvelles pratiques (phase d'apprentissage du programme) qui de plus ne correspondent pas toujours à leur manière d'enseigner habituellement. Cet aspect lié au moment précis du processus d'appropriation d'un programme devrait par conséquent être considéré avec précaution lors de l'évaluation des effets d'un nouvel outil.

C. L'accompagnement de l'enseignant

L'importance de l'accompagnement en tant que chercheur/formateur des enseignants tout au long de l'implémentation d'un nouveau programme constitue également un aspect central à considérer, car il peut affecter positivement ou négativement la mise en œuvre d'un protocole expérimental. Il ne suffirait pas de transmettre des « recettes » ou procédures « toutes faites » à appliquer aveuglément pour qu'un outil soit utilisé avec une fidélité et une qualité d'implémentation élevée. La formation et l'assistance technique constituent un paramètre essentiel à la conduite d'une intervention (Durlak & DuPre, 2008). En effet, il importe que les enseignants soient au clair sur les sources théoriques et empiriques sous-jacentes à l'ensemble de l'intervention proposée et ce afin qu'ils puissent donner du sens aux actions professionnelles qu'ils seront amenés à mettre en œuvre. À titre d'exemple, les concepteurs de certains programmes curriculaires populaires aux États-Unis, comme c'est le cas des « Outils de la pensée » (*Tools of the Mind* ; Bodrova & Leong, 2001), considèrent que leur programme ne peut pas simplement être pris « sur une étagère » et mis en œuvre. Un usage adapté du programme va dépendre du degré de compréhension par les enseignants, dans cet exemple, du cadre de référence théorique historico-culturel qui sous-tend la conception de l'apprentissage et du développement des enfants, mais également des transformations opérées sur les représentations relatives au rôle que l'enseignant va jouer dans le soutien au développement des enfants (Nesbitt & Farran, 2021). L'accompagnement des enseignants devient cependant tout particulièrement délicat lors du passage de la petite échelle à la grande échelle (extension à différents contextes). Bressoux, (2020) insiste alors sur l'importance d'être attentif à ce qu'il nomme « l'encadrement » de l'expérimentation, comprenant les consignes, la manière d'informer, de former les enseignants, d'assurer leur suivi etc. Le nombre de participants impliqués dans l'implémentation du programme qui augmente à grande échelle ainsi que la qualité de l'accompagnement peuvent affecter la qualité de cet encadrement et par conséquent potentiellement générer des résultats décevants.

D. La triade : « vouloir » / « pouvoir » / « savoir »

Nous concluons sur la nécessité de prêter attention aux causes qui peuvent affecter négativement la généralisation d'un programme à grande échelle et qui concerne l'enseignant. Bressoux (2020) propose trois causes potentielles : l'enseignant « ne veut pas », « ne peut pas » et/ou « ne sait pas » implémenter le programme. Lorsque ce dernier « ne veut pas », il importe de questionner différents éléments comme la discordance possible entre le programme et ses connaissances théoriques, son expérience ou sa difficulté à le mettre en œuvre. Pour certains chercheurs, les enseignants devraient pouvoir décider d'adhérer ou non à la « philosophie » d'un programme et de l'implémenter dans sa classe. En d'autres termes, les enseignants devraient bénéficier d'un pouvoir décisionnel dans les outils qu'ils décideront de mettre en œuvre (Nesbitt & Farran, 2021). S'il ne « peut pas », des facteurs

externes, comme la réticence des parents, des élèves, des collègues ou le fait que le programme n'ait pas l'effet escompté sur les élèves, peuvent empêcher la bonne conduite du programme. Le manque d'expérience ou d'outils théoriques, de détails sur la manière d'implémenter l'outil innovant, ne pas bénéficier d'un encadrement de qualité pourraient finalement expliquer les difficultés de l'enseignant à « savoir » implémenter le programme. Ces trois aspects nous offrent des pistes d'analyses intéressantes à utiliser avec les enseignants (par le biais, d'entretiens ou de questionnaire à remplir) dans le cas où un protocole expérimental ne fournit pas les résultats escomptés.

Ces différents constats, non exhaustifs, donnent ou redonnent une place privilégiée au rôle joué par le praticien dans cette démarche scientifique visant à attester de l'efficacité d'une intervention. Toutefois, même si le chercheur et le praticien sont incontestablement indispensables à la création d'un nouveau dispositif ainsi qu'à son évaluation, il n'en demeure pas moins que la place laissée au bénéficiaire de l'intervention (l'enfant dans le contexte scolaire) demeure actuellement peu prise en compte dans la démarche de conception d'outils innovants. Le « *design participatif* » pourrait constituer une piste prometteuse à développer en contexte scolaire, puisqu'il cible l'utilisateur. L'enfant est alors activement impliqué à différents niveaux du processus de conception lui-même. Cette démarche de « partenaire design » dans lequel l'enfant participe de manière égale dans l'élaboration du matériel par exemple est complexe à mettre à œuvre, mais a l'avantage de considérer les besoins, capacités et idées inventives de l'enfant (Valente, Bara & Gentaz, 2018).

Références bibliographiques

- Bara, F. & Gentaz, E. (2011). Haptics in handwriting teaching: the role of perceptual and motor skills. *Human Movement Science*, 30, 745-759.
- Bara, F., Gentaz, E. & Colé, P. (2004). Les effets des entraînements phonologiques et multisensoriels destinés à favoriser l'apprentissage de la lecture chez les jeunes enfants. *Enfance*, 2-3, 145-167.
- Bara, F., Gentaz, E. & Colé, P. (2007). The visuo-haptic and haptic exploration increases the decoding level of children coming from low-socioeconomic status families. *British Journal of Developmental Psychology*, 25, 643-663.
- Bara, F., Gentaz, E., Cole, P. & Sprenger-Charolles, L. (2004). The visuo-haptic and haptic exploration of letters increases the kindergarten-children's understanding of the alphabetic principle. *Cognitive Development*, 19(3), 433-449. doi:10.1016/j.cogdev.2004.05.003
- Bernard, C. (1878). *La science expérimentale*. Paris : Baillière & Fils.
- Blachman, B., Ball, E., Black, R. & Tangel, D. (1994). Kindergarten teachers develop phonemic awareness in low-income inner-city classrooms; Does it makes a difference? *Reading and Writing: An Interdisciplinary Journal*, 6, 1-18.
- Bodrova, E. & Leong, D. (2001). *Tools of the mind : A case study of implementing the vygotskian approach in american early childhood and primary classrooms*. Innodata Monographs 7. Geneva: International Bureau of Education.
- Bodrova, E. & Leong, D. (2012). *Les outils de la pensée*. Québec : PUQ.
- Bressoux, P. (2020). À quelles conditions peut-on déployer à grande échelle les interventions qui visent à améliorer les pratiques enseignantes ? In M. Janosz & B. Galand (Eds.), *Améliorer les pratiques en éducation : Qu'en dit la recherche ?* (pp. 13-22). Louvain : Presses universitaires de Louvain.
- Buyse, V., Sparkman, K. & Wesley, P. (2003). Communities of practice: Connecting what we know with what we do. *Exceptional Children*, 69(3), 263-277. doi:https://doi.org/10.1177/001440290306900301
- Cèbe, S., & Goigoux, R. (2018). Lutter contre les inégalités : Outiller pour former les enseignants. *Recherche & Formation*, 87, 77-96.
- Committee-For-Children. (2012). *Second Step. A violence prevention curriculum*. Seattle, WA: Gallaudet University.
- Dane, A. & Schneider, B. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, 18, 23-45. doi: https://doi.org/10.1016/S0272-7358(97)00043-3
- Deheane, S., Dehaene-Lambert, G., Gentaz, E., Huron, C. & Sprenger-Charolles, L. (2011). *Apprendre à lire. Des sciences cognitives à la salle de classe*. Paris : Odile Jacob.
- Dessus, P. & Gentaz, E. (Eds.) (2006). *Apprentissages et enseignement. Sciences cognitives et Éducation*. Paris : Dunod.

Djubegovic, B. & Guyatt, G. (2017). Progress in evidence-based medicine: A quarter century on. *Lancet*, 22, 415-423.

Durlak, J. & DuPre, E. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3-4), 327-350.

Dusenbury, L., Brannigan, R., Falco, M. & Hansen, W. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*, 18, 237-256.

Eccles, M. & Mittman, B. (2006). Welcome to Implementation Science. *Implementation Science*, 1.

Ehri, L., Nunes, S., Willows, D., Schuster, D., Yaghoub-Zadeh, Z. & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National reading Panel's meta-analysis. *Reading Research Quarterly*, 36, 250-287.

Fisher, R., Smith, K., Finney, S. & Pinder, K. (2014). The importance of implementation fidelity: Data for evaluating program effectiveness. *About Campus*, 19, 28-32. doi:<https://doi.org/10.1002/abc.21171>

Forman, S., Shapiro, E., Coddington, R., Gonzales, J., Reddy, L., Rosenfield, S. & Stoiber, K. (2013). Implementation science and school psychology. *School Psychology Quarterly : The Official Journal of the Division of School Psychology, American Psychological Association*, 28, 77-100.

Gentaz, E. (2018). Du labo à l'école : le délicat passage à l'échelle. *La Recherche*, 539, 42-46.

Gentaz, E. (2019). Les pédagogies alternatives ont déjà transformé l'enseignement. *Cerveau & Psycho*, 116.

Gentaz, E. (2020). Édito - Que peuvent nous apprendre les recherches sur l'évolution de la démarche de la médecine fondée sur des preuves (Evidence-Based Medicine - EBM) ? *A.N.A.E.*, 165, 125-127.

Gentaz, E. (à paraître). *Quels sont les véritables apports des neurosciences aux apprentissages et à l'enseignement ?* Paris : Odile Jacob.

Gentaz, E., Colé, P. & Bara, F. (2003). Évaluation d'entraînements multisensoriels de préparation à la lecture pour les enfants de grande section de maternelle. *L'Année Psychologique*, 104(4), 561-584.

Gentaz, E. & Dessus, P. (Eds.) (2004). *Comprendre les apprentissages. Sciences cognitives et éducation*. Paris : Dunod.

Gentaz, E., Richard, S. & Rivier, C. (2021). Évaluation des effets de la pédagogie Montessori sur le développement psychologique de l'enfant et les apprentissages. In F. Darbellay, Z. Moody & M. Louviot (Eds.), *L'école autrement ? Les pédagogies alternatives en débat* (pp. 205-227). Neuchâtel, Suisse : Alphil - Presses Universitaires Suisse.

Gentaz, E., Sprenger-Charolles, L., Colé, P., Theurel, A., Gurgand, M., Huron, C. & Le Cam, M. (2013). Évaluation quantitative d'un entraînement à la lecture à grande échelle pour des enfants de CP scolarisés en réseaux d'éducation prioritaire : apports et limites. *A.N.A.E.*, 123, 172-181.

Gerstner, J. & Finney, S. (2013). Measuring the implementation fidelity of student affairs programs: A critical component of the outcomes assessment cycle. *Research and Practice in Assessment*, 8, 15-28.

Haynes, R., Straus, S., Glasziou, P. & Richardson, W. (2007). *Médecine fondée sur les faits. Evidence-based medicine*. Paris : Elsevier.

IES (2013). Randomized controlled trials commissioned by the Institute of Education Sciences (IES) since 2002: How many found positive versus weak or no Effects. *Institute of Education Sciences (IES)* - http://coalition4evidence.org/wp-content/uploads/2013/06/IES-Commissioned-RCTs_-_positive-vs-weak-or-null-findings-7-2013.pdf.

Inserm (2007). *Dyslexie, dysorthographe, dyscalculie : bilan des données scientifiques* : Rapport. Paris : Les éditions Inserm, 2007, XV - 842 p. - (Expertise collective).

Inspection générale de l'Éducation nationale. (2011). *L'école maternelle - rapport n°2011-108*. Paris : MEN.

Kalénine, S., Pinet, L. & Gentaz, E. (2011). The visuo-haptic and haptic exploration of geometrical shapes increases their recognition in preschoolers. *International Journal of Behavioral Development*, 35, 18-26.

Klingner, J., Boardman, A. & McMaster, K. (2013). What Does it Take to Scale up and Sustain Evidence-Based Practices? *Exceptional Children*, 79(3), 195-211. doi:<https://doi.org/10.1177/001440291307900205>

Labat, H., Ecalle, J., Baldy, R. & Magnan, A. (2014). How can low-skilled 5-year-old children benefit from multisensory training on the acquisition of the alphabetic principle? *Learning and Individual Differences*, 29, 106-113. doi:<https://doi.org/10.1016/j.lindif.2013.09.016>

Labat, H., Magnan, A. & Ecalle, J. (2011). Effet d'une exploration « multisensorielle séquentielle orientée » sur le développement de la compréhension du principe alphabétique chez les enfants de 5 ans faibles connaisseurs de lettres. *L'Année psychologique*, 111, 641-671.

Landry, S. (2014). *Favoriser le développement de la pensée sociale par l'étayage du jeu symbolique chez des enfants qui fréquentent la maternelle cinq ans (Thèse de doctorat)*. Laval : Université de Laval. <https://corpus.ulaval.ca/jspui/handle/20.500.11794/24789>.

Laroche, P. (2015). *La méta-analyse. Méthodes et applications en sciences sociales*. Louvain la Neuve : De Boeck.

Malsert, J., Theurel, A. & Gentaz, E. (2013). Évaluation des effets d'un programme « Atelier cognitif (AC) : comprendre le fonctionnement du cerveau pour mieux apprendre » sur les performances d'élèves de seconde générale. *A.N.A.E.*, 123, 120-126.

Mayer, R. (1992). Cognition and instruction: Their historic meeting within educational psychology. *Journal of Educational Psychology*, 84, 405-412.

National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. Washington, D.C.

Nesbitt, K. & Farran, D. (2021). Effects of prekindergarten curricula: Tools of the mind as a case study. *Monographs of the Society for Research in Child Development*, 86(1), 7-119.

Odom, S., Hall, L. & Suhrheinrich, J. (2020). Implementation science, behavior analysis, and supporting evidence-based practices for individuals with autism. *European Journal of Behavior Analysis*, 21, 55-73.

Pinet, L. & Gentaz, E. (2008). Évaluation d'entraînements multisensoriels de préparation à la reconnaissance de figures géométriques planes chez les enfants de 5 ans : étude de la contribution du système haptique manuel. *Revue française de pédagogie*, 162, 29-44.

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A. & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health*, 38(2), 65-76. doi: <https://doi.org/10.1007/s10488-010-0319-7>

Richard, S., Clerc-Georgy, A., Baud-Bovy, G. & Gentaz, E. (2021). The effects of a 'pretend play-based training' designed to promote the development of emotion comprehension, emotion regulation, and prosocial behaviour in 5- to 6-year-old Swiss children. *British Journal of Psychology*, 112, 690–719.

Richard, S., Clerc-Georgy, A. & Gentaz, E. (soumis). The effects of a 'pretend play-based training' in schools designed to enhance the development of some socio-emotional and pretend play competences in 5- to 6-year-old Swiss children: a large scale study.

Richard, S., Gay, P., Clerc-Georgy, A. & Gentaz, E. (2019). Évaluation d'un entraînement basé sur le jeu de faire semblant destiné à favoriser le développement des compétences socio-émotionnelles chez les enfants de cinq ans : Étude exploratoire. *L'Année Psychologique*, 119, 291-332.

Richard, S. & Gentaz, E. (2020). Le jeu de faire semblant favorise-t-il le développement des compétences socio-émotionnelles ? *A.N.A.E.*, 32, 172-182.

Slavin, R. (2020). How evidence-based reform will transform research and practice in education. *Educational Psychologist*, 55(1), 21-31. doi: <https://doi.org/10.1080/00461520.2019.1611432>

Theurel, A. & Gentaz, E. (2015). Entraîner les compétences émotionnelles à l'école. *A.N.A.E.*, 139, 545-555.

Valente, D., Bara, F. & Gentaz, E. (2018). *Un guide pour concevoir des livres multisensoriels accessibles à tous avec la méthode du design participatif. Exemple d'un livre conçu avec les enfants en situation de handicap visuel*. Genève, Suisse : Université de Genève.

Vygotsky, L. S. (1933/2016). Play and its role in the mental development of the child. *International Research in Early Childhood Education*, 7(2), 3-25.

White, R., Thibodeau-Nielsen, R., Palermo, P. & Mikulski, A. (2021). Engagement in social pretend play predicts preschoolers' executive function gains across the school year. *Early Childhood Research Quarterly*, 56, 103-113.

Zorman, M., Bressoux, P., Bianco, M., Lequette, C., Pouget, G. & Pourchet, M. (2015). « PARLER » : un dispositif pour prévenir les difficultés scolaires. *Revue française de pédagogie*, 193, 56-76.



le **cnam**
Cnesco

Centre national d'étude des systèmes scolaires

CENTRE NATIONAL D'ÉTUDE DES SYSTÈMES SCOLAIRES
CONSERVATOIRE NATIONAL DES ARTS ET MÉTIERS

41 rue Gay-Lussac - 75005 PARIS

06 98 51 82 75 - cnesco@lecnam.net

www.cnesco.fr



RÉSEAU CANOPÉ

4 avenue du Futuroscope Téléport 1

CS 80158 86961 FUTUROSOCPE

05 16 01 76 26 / www.reseau-canope.fr

FRANCE
ÉDUCATION
INTERNATIONAL

FRANCE ÉDUCATION INTERNATIONAL

1 avenue Léon Journault - 92318 SÈVRES

01 45 07 60 00

www.france-education-international.fr



UNIVERSITÉ
DE GENÈVE

FACULTÉ DE PSYCHOLOGIE
ET DES SCIENCES DE L'ÉDUCATION

UNIVERSITÉ DE GENÈVE

24 rue du Général Dufour - 1211 GENÈVE 4

+41 (0)22 379 71 11

www.unige.ch

RETROUVEZ LES DERNIÈRES ACTUALITÉS DU CNESCO :



www.cnesco.fr



[Cnesco](#)



[Cnesco](#)



[Cnesco](#)