

Référence chapitre :

Galand, B., Baudoin, N., & Tolmatcheff, C. (2023). Avoir un impact : pourquoi mesurer l'implantation d'une intervention est crucial. Dans T. Coppe, A. Baye, & B. Galand (Dir.), *Transformer les pratiques en éducation : Quelles recherches pour quels apports ?* (pp. 69-84). Presses universitaires de Louvain.

Référence ouvrage :

Coppe, T., Baye, A., & Galand, B. (Dir.). (2023). *Transformer les pratiques en éducation : Quelles recherches pour quels apports ?* Presses universitaires de Louvain.

Avoir un impact : pourquoi mesurer l'implantation d'une intervention est crucial.

Benoit Galand, Noémie Baudoin, Chloé Tolmatcheff
Université catholique de Louvain

“The bridge between a promising idea and the impact on students is implementation, but innovations are seldom implemented as intended.” (Berman & McLaughlin, 1979, p. 349)

“The realization of policy in practice thus depends on the fit between capabilities that support implementation and aims.” (Cohen et al., 2007)

Ce chapitre s'interroge sur la manière dont des recherches scientifiques pourraient aider à développer des actions éducatives à large échelle ayant un impact positif sur leurs bénéficiaires. Dans un premier temps, le rôle de la mise en œuvre concrète d'une action planifiée est discuté, que ce soit dans la pratique ou en recherche. Une réflexion sur les étapes du passage d'une connaissance théorique à un impact dans la vie réelle est ensuite proposée. Suite à cela, trois approches complémentaires de la mesure de l'implantation sont présentées, puis nous évoquons plus en détail le débat entre fidélité et adaptation. Enfin, nous revenons sur les raisons pour lesquelles il est si difficile d'observer des effets positifs dans les études d'impact.

D'une bonne idée à un effet bénéfique, le rôle de l'implantation

Que ce soit dans le cadre d'un accompagnement individualisé, d'une intervention auprès d'un groupe ou d'un projet d'école, la plupart des professionnels de l'éducation (enseignants, éducateurs, directions, psychologues, etc.) sont, à un moment ou un autre, confrontés à la question de la suite à donner à leur action : arrêter, poursuivre, renforcer ou changer de stratégie ? Cette décision se fonde généralement sur une forme d'évaluation de l'action planifiée : dans quelle mesure le ou les objectifs fixés ont-ils été atteints ? Cependant, la progression vers les objectifs n'est pas le seul élément à prendre en considération dans ce type de décision. Il est également crucial de se demander dans quelle mesure l'action planifiée a réellement été mise en œuvre (Noell & Gansle, 2006). En effet, comment juger de « l'efficacité » d'une action et de la suite à lui donner si on ignore si elle a bien été mise en place auprès des bénéficiaires visés ? De nombreux spécialistes insistent donc sur l'importance de prendre en compte « l'intégrité de traitement », et pas seulement les

« résultats », dans les décisions que prennent les professionnels quant à l'évolution de leurs interventions (Schulte et al., 2009). Évaluer l'intégrité de traitement apparaît comme un élément essentiel pour apprendre de sa pratique et ajuster les interventions de manière pertinente. Or, plusieurs études indiquent que, même si les professionnels reconnaissent l'importance d'évaluer le degré d'implantation des actions planifiées (intégrité de traitement), ils omettent largement de le faire en pratique (Cochrane & Laux, 2008). Il semble que les professionnels passent plus de temps à réfléchir à ce qu'il faudrait faire qu'à documenter ce qui a réellement été fait.

Les chercheurs sont confrontés à des questions similaires lorsqu'ils tentent d'évaluer les effets d'interventions ou d'innovations censées apporter des effets bénéfiques. Quels que soient les résultats observés concernant les objectifs visés, il s'agit de savoir si ces résultats peuvent être reliés à l'innovation ou l'intervention imaginée. Pour ce faire, il est nécessaire de disposer d'informations sur le degré de mise en œuvre de cette intervention ou innovation : Combien d'activités prévues ont-elles été réalisées avec les bénéficiaires et de quelle manière ? Dans quelle mesure les intervenants ont-ils modifié les consignes ou ajouté des éléments inattendus ? De nombreux chercheurs parlent de fidélité d'implantation pour désigner ce type d'informations, qui sont cruciales pour assurer la validité interne du dispositif d'évaluation et pouvoir interpréter les résultats (peut-on réellement attribuer les résultats à l'intervention imaginée ?). C'est d'autant plus vrai lorsque l'intervention ne produit pas les résultats escomptés, car cela permet de savoir si ceux-ci peuvent s'expliquer par des difficultés de mise en œuvre ou s'il faut remettre en question les fondements de l'intervention elle-même (Nelson et al., 2012). Or, il n'est pas rare qu'une intervention, même fondée sur des modèles théoriques bien établis et prometteuse à petite échelle, n'ait pas les effets escomptés une fois diffusée à large échelle (par exemple, Gentaz et al., 2013 ; Munier et al., 2021 ; Spence & Shortt, 2007 ; West & O'Neal, 2004). Recueillir des informations concernant l'implantation peut aider à identifier les composantes essentielles des innovations ou interventions et à améliorer la diffusion et la durabilité de celles-ci (Latessa, 2018). Noell et al. (2009) estiment qu'énormément d'attention a été portée aux questions éthiques et théoriques dans les recherches sur les changements de pratiques, mais que les questions pragmatiques, liées au soutien à la mise en œuvre des changements, ont été négligées. Ces questions d'implantation apparaissent pourtant cruciales pour le passage de la recherche à la pratique.

Ainsi, la question de la mise en œuvre ou de l'implantation s'avère être un enjeu important, autant pour la pratique des professionnels de l'éducation que pour celle des chercheurs en éducation. C'est aussi vrai dans de nombreux domaines connexes à l'éducation, comme la prévention, la promotion de la santé, la justice (Schulte et al., 2009), même si ce texte se concentre sur le milieu scolaire.

De la théorie à la pratique : Les défis de la recherche translationnelle

La vision de nombreux décideurs et de nombreux chercheurs concernant la diffusion d'innovation ou de changement en éducation semble très proche d'un modèle « recherche, développement et diffusion » (Blakely et al., 1987). Selon cette conception, l'élaboration de nouvelles idées, ou la découverte de nouvelles connaissances, peut servir au développement de nouvelles pratiques ou de nouveaux outils, plus ou moins testés à petite échelle, qui sont ensuite diffusés au sein du système éducatif, de santé, judiciaire, etc. Ce modèle a été largement critiqué pour son simplisme et sa linéarité. Une des principales critiques porte sur le fait que le modèle laisse penser que les acteurs dont les pratiques sont visées par les

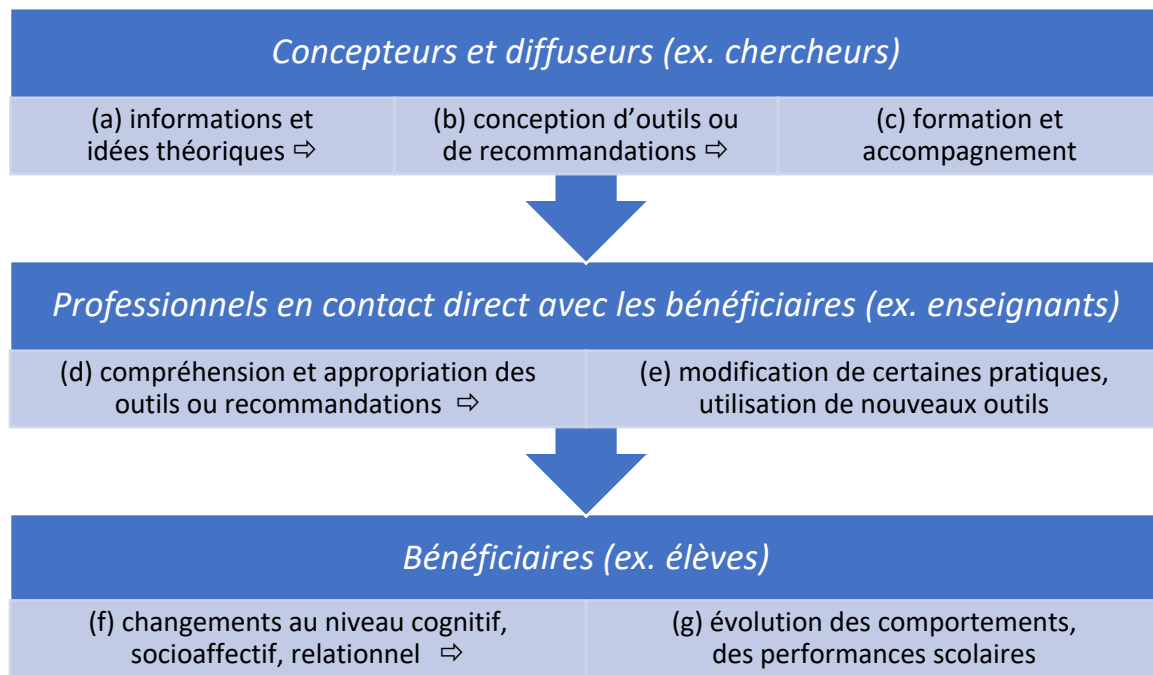
innovations ne seraient que des récepteurs passifs, alors que de nombreuses études montrent que les acteurs filtrent et transforment activement les outils ou préconisations qui leur sont adressées (Blakely et al., 1987). Cette critique renvoie entre autres au débat entre fidélité et adaptation, sur lequel nous reviendrons plus tard. L'introduction de changements ou d'innovations dans les systèmes s'avère effectivement plus compliquée que ne le laisse penser ce modèle (Bressoux, 2020 ; Lipsey, 2020), en éducation comme ailleurs (Duflo, 2020 ; Prasad & Cifu, 2015).

Ce constat a amené de nombreux spécialistes des interventions à poser une distinction entre études d'efficacité (*efficacy*) et études d'impact (*effectiveness*) (Gottfredson et al., 2015 ; O'Donnell, 2008). Les études d'efficacité consistent à tester les effets d'une intervention dans des conditions optimales et les plus contrôlées possibles, pour établir de la manière la plus valide possible que l'intervention est bien capable de produire les effets attendus sur le terrain. Il s'agit souvent d'études menées par les concepteurs de l'intervention eux-mêmes, auprès d'un nombre relativement restreint de participants généralement volontaires, avec un accompagnement habituellement soutenu. Les études d'impact consistent quant à elles à tester les effets d'une intervention dans des conditions les plus proches possible des conditions habituelles, afin d'examiner si l'intervention peut produire des effets une fois diffusée à large échelle (Nault Brière, 2017). Il s'agit par conséquent d'études menées avec plusieurs intervenants (formateurs, coaches...), sur plusieurs sites, auprès d'un grand nombre de participants pas nécessairement volontaires, avec un encadrement et des ressources souvent plus restreints. D'autres chercheurs parlent d'études de démonstration et d'études de routine pour désigner respectivement les études d'efficacité et les études d'impact (Wampold & Imel, 2015 ; Wilson et al., 2003).

Selon tous ces chercheurs, le passage d'une « découverte » scientifique à une application pratique nécessite non seulement de montrer qu'un résultat « de laboratoire » est capable de produire des effets concrets (étude d'efficacité), mais aussi que ces effets se vérifient une fois l'innovation « lâchée dans la nature » (étude d'impact). Dans les faits, de nombreuses « innovations » s'affichant comme « s'appuyant sur des preuves scientifiques » sont diffusées en l'absence d'études d'efficacité (De Bruyckere et al., 2015 ; Tolmatcheff et al., 2020). De plus, comme évoqué précédemment, de nombreuses interventions disposant d'études d'efficacité n'ont pas d'effets détectables dans des études d'impact (Bressoux, 2020 ; Kim, 2019). Une étude de Hulleman et Cordray (2009) illustre bien ce problème. Elle montre qu'une intervention qui s'avère efficace pour favoriser la motivation des élèves lorsqu'elle est réalisée par des chercheurs en conditions contrôlées s'avère inefficace quand elle doit être menée par des enseignants. Les résultats montrent aussi que la fidélité d'intervention est beaucoup plus faible de la part des enseignants. Le passage d'une connaissance scientifique à son application concrète sur le terrain se révèle par conséquent plus complexe que beaucoup ne l'imaginent. La figure 1 présente une tentative de schématiser les étapes de ce passage (en élaborant notamment à partir de Rimm-Kaufman et al., 2014).

Figure 1.

Étapes du passage d'une connaissance théorique à un effet pour les bénéficiaires



Dans les faits, de nombreux concepteurs se fondent sur leur expérience, leurs réflexions, leurs échanges et leurs lecteurs pour concevoir des interventions ou des innovations. Les chercheurs s'efforcent pour leur part de s'appuyer sur des données empiriques et des modèles théoriques partagés au sein de la communauté scientifique. Même si elle reste très linéaire, la figure 1 illustre combien les étapes allant d'une connaissance scientifique à un effet positif éventuel pour les bénéficiaires sont nombreuses et que chaque étape soulève des défis spécifiques qui peuvent hypothéquer les bénéfices escomptés (Spoth et al., 2013). Les idées de départ peuvent être erronées, leur transformation en pistes pratiques peut être inadéquate, les méthodes ou supports de diffusion peuvent s'avérer inappropriés, les professionnels peuvent mal interpréter ou être en désaccord avec ce qui est proposé, des contraintes organisationnelles peuvent les empêcher de mettre en œuvre les changements attendus, les bénéficiaires peuvent percevoir ces changements d'une manière inattendue, etc. Les enjeux relatifs à la mesure de l'implantation, qui font l'objet de ce texte, concernent principalement les étapes (c), (d) et (e) de la figure 1.

À ce sujet, de nombreux chercheurs parlent d'intégrité de traitement ou de fidélité d'implantation. Gresham et al. (2000) définissent l'intégrité de traitement comme le degré avec lequel le traitement est mis en œuvre comme prévu. Mowbray et al. (2003) définissent, quant à eux, la fidélité comme le degré avec lequel la mise en œuvre d'une intervention correspond au protocole ou au modèle conçu initialement. Carroll et al. (2007) estiment pour leur part que la fidélité d'implantation est le degré avec lequel une intervention ou un programme est déployé comme prévu. O'Donnell (2008) a recensé de nombreuses autres définitions de la fidélité d'implantation et constate que les définitions et les mesures ne sont pas cohérentes d'une étude à l'autre. Il n'empêche que toutes ces définitions tournent autour de la thématique de la correspondance entre les actions annoncées, planifiées, et ce qui est effectivement réalisé, et que les termes d'intégrité de traitement et de fidélité d'implantation sont largement synonymes. Dans le cadre de ce texte, retenons que mesurer l'implantation revient à récolter des informations sur ce qui a été fait concrètement par rapport à ce qu'on s'était engagé à faire ou qu'on avait accepté de faire (Dellisse & Galand, 2020).

Trois façons de s'intéresser à la question de la mesure d'implantation peuvent être distinguées dans la littérature scientifique, correspondant globalement à une évolution dans les préoccupations des chercheurs. La première consiste à mesurer l'implantation juste pour s'assurer d'un niveau d'application suffisant de l'intervention. La seconde consiste à mettre en relation le niveau ou la qualité d'implantation avec les résultats de l'intervention. La troisième consiste à analyser les facteurs pouvant expliquer le degré d'implantation d'une intervention. Chacune de ces approches est détaillée ci-dessous.

Mesurer l'implantation pour s'assurer de l'application réelle de l'intervention

Une première approche consiste à mesurer l'implantation simplement pour estimer si l'intervention ou l'innovation planifiée a bien été délivrée aux bénéficiaires. L'enjeu est ici surtout une question de validité : ce qu'on évalue est-il réellement l'action prévue ? Par exemple, Clements et Sarama (2008) ont utilisé un outil d'observation pour mesurer l'implantation d'un nouveau curriculum de mathématiques en maternelle et s'assurer que celui-ci avait été utilisé, avant d'évaluer les effets de ce curriculum sur les progrès des élèves en comparaison avec un groupe n'ayant pas accès à ce curriculum. Mais cette pratique est-elle courante ?

Passant en revue 181 études expérimentales concernant une intervention comportementale et publiées entre 1980-1990, Gresham et al. (1993) constatent que seules 14,4 % d'entre elles mesurent de façon systématique et rapportent le niveau de fidélité d'implantation. De plus, seule une minorité d'études expliquent de façon opérationnelle en quoi consiste l'intervention. Analysant les interventions auprès d'élèves ayant des difficultés d'apprentissage publiées dans trois des principaux journaux dans ce domaine, Gresham et al. (2000) observent que seules 18,5 % des études mesurent l'intégrité de traitement. Ils notent la présence d'un double standard méthodologique dans les mesures : celles des variables dépendantes ou de sortie font l'objet de nombreux critères de qualité (fiabilité, validité), tandis que celles des variables indépendantes (traitement, intervention ou innovation) font l'objet de peu d'exigences. Domitrovich et Greenberg (2000) dressent un constat similaire dans le champ de la santé mentale, indiquant que la plupart des programmes présentés comme efficaces ne mesurent pas la fidélité de la mise en œuvre. Examinant les études d'intervention publiées dans le *Journal of Applied Behavior Analysis* de 1991 et 2005, McIntyre et al. (2007) notent que seules 30 % d'entre elles rapportent des données d'intégrité de traitement. Ils constatent peu d'amélioration au fil des ans, sauf concernant la définition opérationnelle du traitement.

Les études qui ont mesuré l'implantation montrent que celle-ci est généralement loin d'être parfaite (Durlak & Dupré, 2008). Ainsi, la plupart des enseignants ne mettent pas en œuvre l'ensemble de l'intervention ou du curriculum proposé (Dusenbury et al., 2003) et y apportent des modifications (Datnow & Castellano, 2000), et la majorité des équipes éducatives n'appliquent pas complètement ou détournent partiellement les démarches ou les innovations proposées (De Smet et al., 2017 ; Janosz et al., 2011).

Au vu de ces résultats et comme l'indique Latessa (2018), il est possible qu'une part de ce qui a été évalué comme n'étant pas efficace ne se soit en fait tout simplement pas produit. Néanmoins, plusieurs études suggèrent que mesurer la fidélité d'implantation serait désormais une pratique plus courante dans la recherche en éducation. La majorité des études plus récentes examinées par Hill et Erikson (2019) concernant des interventions éducatives, rapportent en effet certaines mesures d'implantation. De même, de Leew et al. (2020) ont

passé en revue les études à propos des interventions en classe qui visent à faciliter la participation sociale des élèves présentant des problèmes socioémotionnels ou comportementaux. Les sept études qu'ils ont identifiées rapportent toutes des informations sur la fidélité d'implantation, mais ne donnent en revanche pas de détails sur les composantes cruciales des interventions et le lien entre fidélité d'implantation et efficacité de l'intervention. Ces dernières études suggèrent que si le fait de mesurer la fidélité d'implantation devient peut-être plus courant, d'autres enjeux subsistent.

Étudier les liens entre degré d'implantation et effets

Une deuxième approche de l'implantation consiste à examiner si le degré de fidélité est relié aux effets observés. L'enjeu est cette fois de tester si une plus grande fidélité d'implantation conduit à des effets plus prononcés, ce qui soutiendrait la pertinence de l'intervention et l'idée que les effets sont causés par l'intervention. Une telle approche peut aussi contribuer à déterminer s'il existe des seuils de mise en œuvre minimaux ou maximaux afin d'obtenir des effets de l'intervention. Enfin, si différentes composantes de l'intervention ou de l'innovation ont fait l'objet d'une mesure spécifique d'implantation, il est possible d'essayer d'identifier la ou les composantes qui contribuent le plus aux effets observés, qui seraient en quelque sorte les « ingrédients actifs » (Latessa, 2018 ; Lipsey, 2020).

Une association entre implantation et effets

Un grand nombre d'études ont trouvé une relation positive, d'ampleur faible à modérée, entre degré de fidélité et taille des effets, que ce soit concernant la motivation (Hulleman & Cordray, 2009), la lecture (Dellisse et al., 2021), les mathématiques (Hamilton et al., 2003), les sciences (Hill & Erikson, 2019), le climat scolaire (Janosz et al., 2011), les difficultés de comportement (Gresham et al., 1993), les comportements antisociaux (Goense et al., 2016), le harcèlement (Haataja et al., 2014), la récidive d'adolescents délinquants (Baglivio et al., 2018), l'abus de substances (Dusenbury et al., 2003), les difficultés émotionnelles (Askell-Williams et al., 2013), etc.

Évaluant les effets de *Comprehensive School Reforms* (CSR), une politique menée aux États-Unis incitant les écoles à mettre en œuvre des programmes innovants en vue d'accroître la réussite des élèves, Zhang et al. (2006) constatent que les élèves des écoles ayant adopté des CSR n'obtiennent pas de meilleurs résultats que les autres. En revanche, des analyses au niveau établissement montrent une relation positive entre le niveau moyen de progression des élèves en mathématiques et langues, d'une part, et le type de CSR choisi, le niveau d'implantation de celui-ci et la durée d'implantation, d'autre part. Dans une méta-analyse concernant les effets de formations des élèves aux stratégies d'apprentissage sur leurs résultats, de Boer et al. (2014) trouvent aussi un lien entre la durée des sessions de formation et les résultats. Inversement, des évaluations de projets de prévention des violences à l'école (Galand, 2011) ou du harcèlement entre élèves (Ttofi & Farrington, 2011) n'ont pas pu mettre en évidence de relation entre la durée de ces projets et leurs effets sur les élèves. De leur côté, Giles et al. (2008) ont observé l'utilisation par les enseignants de compétences d'enseignement interactif promues par un projet de prévention des addictions. Ils constatent que l'utilisation de ces compétences n'est que marginalement liée à la consommation par les élèves, et ce, pour une seule des substances investiguées. Au terme d'une revue de la littérature sur les interventions concernant le curriculum de l'enseignement obligatoire, O'Donnell (2008) estime qu'il y a trop peu d'études pour guider les chercheurs concernant la

façon de mesurer la fidélité d'implantation et la façon de la relier aux résultats. Goulet et al. (2018) constatent, quant à eux, que les effets de la fidélité à différentes composantes d'un programme de prévention du décrochage sont différents en primaire et en secondaire, et selon la formation initiale des intervenants.

Face à ces résultats parfois incohérents, Durlak et Dupré (2008) ont réalisé une méta-analyse sur plus de 500 études quantitatives. Les résultats de cette méta-analyse soutiennent clairement l'existence d'une relation positive entre le niveau d'implantation des projets de promotion ou de prévention et les résultats obtenus. Ce type de résultats empiriques a amené des chercheurs à approfondir la notion de fidélité d'implantation et à distinguer plusieurs dimensions de l'implantation.

Différentes dimensions et mesures de l'implantation

Avec le développement de l'intérêt pour l'implantation, plusieurs chercheurs ont proposé d'identifier des sous-dimensions de l'implantation. Dusenbury et al. (2003) ont, par exemple, passé en revue les études sur la prévention d'abus de substances et identifié cinq dimensions de la fidélité : l'adhérence, le dosage, la qualité du déploiement, la réceptivité des participants et la différenciation de l'intervention. Carroll et al. (2007) évoquent l'adhérence au contenu, l'ampleur couverte, la fréquence et la durée comme dimensions distinctes. Proctor et al. (2011) proposent, quant à eux, d'évaluer les aspects suivants de l'implantation : l'acceptabilité, l'adoption, la faisabilité, la fidélité, le coût, la pénétration et la durabilité. Bref, il n'y a actuellement pas de consensus scientifique sur les dimensions clés de l'implantation, même s'il semble y avoir un accord sur le fait qu'il s'agit d'un concept multidimensionnel (O'Donnell, 2008). En outre, Berkel et al. (2011) soulignent que peu d'études investiguent plus d'une dimension de l'implantation à la fois et qu'on manque de cadres théoriques pour guider les recherches sur les effets de l'implantation. Il est par conséquent impossible aujourd'hui de savoir dans quelle mesure certaines dimensions de l'implantation seraient plus cruciales que d'autres.

Cet état de la situation complique aussi le développement de mesures d'implantation, leur validation et leur utilisation cohérente à travers différentes études (O'Donnell, 2008 ; Proctor et al., 2011). Certaines études se basent sur des observations, d'autres sur des entretiens ou des questionnaires, d'autres encore croisent différentes sources d'information (Askill-Williams et al., 2013 ; Dusenbury et al., 2003). Certains chercheurs ont proposé des outils et des procédures de validation (Buckley et al., 2017 ; Mowbray et al., 2003 ; Nelson et al., 2012), mais les études empiriques sont encore lacunaires. Une étude indique que les personnes qui mettent en œuvre les interventions tendent à rapporter un niveau d'implantation plus élevé que des observateurs indépendants, même si les évaluations sont corrélées entre différents évaluateurs (Lillehoj et al., 2004). Cette étude montrait aussi que l'implantation rapportée par les observateurs était associée à plusieurs effets liés à l'usage de substances parmi les jeunes, ce qui n'était pas le cas de l'implantation rapportée par les professionnels chargés de l'intervention. Ces résultats suggèrent que toutes les mesures d'implantation ne se valent pas et que les questions de mesure constituent un enjeu important pour le développement de ce champ de recherche (Haataja et al., 2014).

Identifier et comprendre les facteurs qui influencent l'implantation

Étant donné la difficulté à observer des effets positifs dans les études d'impact et le lien positif récurrent entre niveau d'implantation et taille d'effet, un nombre croissant d'études se sont penchées sur les facteurs qui pourraient expliquer ou prédire l'implantation. L'enjeu de cette troisième approche de la mesure de l'implantation est de contribuer à une théorie du changement et de dégager des pistes pour améliorer la mise en œuvre d'interventions ou d'innovations.

Au niveau individuel, plusieurs études ont trouvé des liens entre les croyances, les attitudes et l'expérience préalable que les enseignants ont de l'intervention d'une part, et le niveau d'implantation d'autre part (Gregoire, 2003 ; Holdsworth & Maynes, 2017 ; Lillehoj et al., 2004). Quelques études indiquent que l'implantation d'activités d'apprentissage varie peu d'un élève à l'autre, mais beaucoup d'un enseignant à l'autre (Hulleman & Cordray, 2009 ; Rowan et al., 2004), et que ces variations entre enseignants sont très peu fonction de leur genre, de leur ancienneté, de leur formation initiale ou de la composition du groupe d'élèves (Datnow & Castellano, 2000 ; Rowan et al., 2004). Une autre étude observe que la fidélité d'implantation est plus élevée chez les enseignants qui perçoivent davantage que le programme proposé demande des changements de pratiques (Anderson, 2017). Au niveau de l'intervention elle-même, Rowan et Miller (2007) observent qu'une plus grande standardisation des activités entraîne davantage de changement dans les pratiques d'enseignement. Hill et Erikson (2019) constatent également que la mise à disposition de matériel détaillé prédit davantage de mise en œuvre de nouveaux curricula. Au niveau établissement, Gregory et al. (2007) ont observé que la collaboration perçue par les enseignants au sein de l'équipe et avec les élèves prédit l'implantation d'un programme de prévention des violences, tandis que le leadership administratif de la direction prédit l'évolution de l'implantation sur trois ans. Des problèmes de regroupement d'élèves et de planning peuvent aussi entraîner des difficultés de mise en œuvre (Klingner et al., 2006).

La plupart des publications s'accordent sur le fait que des facteurs à différents niveaux sont à prendre en compte pour comprendre l'implantation, mais elles ne s'accordent pas sur les mêmes niveaux et facteurs (Dellisse & Galand, 2020). Les études analysées par Dusenbury et al. (2003) suggèrent par exemple que la formation seule n'est pas suffisante pour assurer la fidélité d'implantation. Les facteurs qui ressortent, outre la formation des enseignants, sont les caractéristiques du programme, des enseignants et des organisations. Mihalic et Irwin (2003) ont collecté des informations sur les difficultés d'implantation sur 42 sites mettant en place huit projets différents de prévention des violences (qui avaient des résultats prometteurs dans des études d'efficacité). Ces difficultés étaient nombreuses, les plus importantes étant apparemment liées à la qualité de l'accompagnement, aux caractéristiques du programme, à la stabilité des équipes et au soutien de la communauté. Les résultats d'une vaste enquête, menée auprès de centaines de directions, indiquent des relations entre l'implantation et le processus de développement local, l'intégration dans les procédures existantes de l'établissement, les capacités organisationnelles, le soutien de la direction et la standardisation du projet (Payne et al., 2006). Une autre étude montre que la perception du soutien de la direction, le peu d'expérience des enseignants en littératie et la formation des coaches au programme visé, sont associés à la participation à des séances de coaching concernant la littératie (Matsumura et al., 2010). Dans leur revue de la littérature, Domitrovich et al. (2008) identifient des facteurs au niveau macro, au niveau des écoles et au niveau individuel, alors que Durlak et Dupré (2008) identifient des variables reliées aux communautés, aux fournisseurs et aux innovations, ainsi qu'au fonctionnement des organisations mettant en œuvre les innovations, et aux formations et à l'accompagnement fournis en soutien à la mise en œuvre.

Au total, il n'est pas facile de déterminer l'importance de différents facteurs pour l'implantation, peu d'études ayant directement comparé des facteurs entre eux de façon quantitative (Spoth et al., 2013) et les études concernant les facteurs de durabilité des interventions en sont encore au stade embryonnaire (Wiltsey Stirman et al., 2012). Une étude de Dellisse et al. (2020) a évalué dans quelle mesure la fidélité d'implantation (rapportée par les élèves) d'un programme de compréhension à la lecture était liée à la perception, par les enseignants, de facteurs relatifs à leur établissement, à eux-mêmes ou au programme. Les résultats indiquent que seuls certains facteurs personnels (par exemple, le sentiment d'efficacité) et relatifs au programme lui-même (par exemple, la clarté des consignes) sont associés à la fidélité d'implantation. Si de nombreuses stratégies ont été proposées pour favoriser l'implantation d'innovations ou d'interventions (Powell et al., 2015), très peu d'entre elles ont été évaluées empiriquement. Dans une des rares études de ce type, Noell et al. (2005) ont comparé les effets de trois stratégies sur l'intégrité de traitement d'enseignants accompagnés par des psychologues scolaires dans le suivi d'un élève : un entretien hebdomadaire, un entretien avec une emphase sur la mise en place du traitement, et des informations sur l'évolution des comportements de l'élève (feedbacks de performance). Les feedbacks de performance amènent davantage d'intégrité de traitement et d'amélioration des comportements des élèves. Aucun lien significatif entre l'acceptabilité du traitement perçue par l'enseignant et le degré d'implantation n'a été observé. Ces études fournissent des indications très intéressantes, mais davantage de recherches sont nécessaires pour développer une base de connaissances solide sur ces questions.

Le débat entre fidélité et adaptation

Le constat, ancien et récurrent, que les professionnels n'implantent pas complètement les interventions ou les innovations et les transforment, nourrit un débat tout aussi ancien et récurrent entre les tenants de la fidélité et les tenants de l'adaptation (Blakely et al., 1987 ; Dusenbury et al., 2003), débat qui a des répercussions pour les trois approches de la mesure d'implantation présentée ci-dessus. D'un côté, les tenants de la fidélité estiment que la conformité au protocole ou aux démarches développées par les concepteurs (souvent considérés comme des experts), est essentielle à l'efficacité de l'intervention. Dans cette perspective, l'objectif est alors de trouver des moyens pour renforcer la fidélité et limiter les adaptations locales. D'un autre côté, les tenants de l'adaptation estiment que l'appropriation par les professionnels et l'adéquation aux contextes locaux et à différents publics sont essentiels à l'efficacité de l'intervention. Dans cette perspective, les transformations et les ajouts par les acteurs locaux sont plutôt à encourager, du moins dans une certaine mesure, et une approche trop rigide de la fidélité est vue comme une entrave.

Vers un équilibre ou vers d'autres modèles ?

En réalité, la plupart des chercheurs mettant en place des interventions se situent quelque part entre ces deux positions extrêmes, reconnaissant que la question est sans doute de trouver un équilibre entre fidélité et adaptation (Forehand et al., 2010). Certains ont proposé de réfléchir au rôle de la théorie dans cet équilibre (Mowbray et al., 2003), d'autres à des façons d'aider les professionnels à appliquer une intervention de manière flexible sans la dénaturer complètement (Mazzucchelli & Sanders, 2010) et d'autres encore à des modèles d'intervention qui combinent fidélité et adaptation (Webster-Stratton et al., 2011). Il est également possible que l'importance de la fidélité varie selon la nature des interventions ou innovations (par exemple, un protocole d'accompagnement individualisé, un curriculum dans

une matière scolaire ou une stratégie d'action au niveau d'un établissement) (O'Donnell, 2008). Cependant, ce débat est rarement appuyé sur des arguments empiriques solides. À cet égard, une étude qualitative de Datnow et Castellano (2000) montre que le niveau d'adhésion des enseignants à une intervention n'est pas nécessairement lié à leur niveau de fidélité dans l'implantation de celle-ci. Beaucoup d'enseignants soutiennent la mise en place de cette intervention, car ils sont convaincus qu'elle est bénéfique pour les élèves, alors que, dans le même temps, ils ont le sentiment que cette intervention contraint leur autonomie et leur créativité. Dans une étude déjà ancienne, Blakely et al. (1987) ont visité 70 sites mettant en place sept interventions différentes. Ils observent qu'un degré plus élevé de fidélité d'implantation est associé à un impact plus fort. Ils observent également que le degré de réinvention locale de l'intervention n'est pas lié à l'efficacité tandis que des ajouts faits localement sont associés à des effets plus prononcés. Plus récemment, Quinn et Kim (2017) ont testé une séquence basée sur l'étayage : une stratégie focalisée sur la fidélité suivie l'année d'après par une stratégie d'adaptation structurée. Dans leur étude, la stratégie focalisée sur la fidélité suscite davantage de changement de pratiques (implantation) et d'amélioration des apprentissages des élèves (efficacité) chez les enseignants n'ayant pas l'expérience de l'intervention, tandis que la stratégie d'adaptation structurée entraîne davantage d'implantation et d'efficacité chez les enseignants déjà familiarisés avec l'intervention. Davantage de recherches de ce type seraient utiles pour faire avancer le débat entre fidélité et adaptation sur des bases renouvelées.

Une ligne de démarcation entre chercheurs reste néanmoins de savoir si cette flexibilité d'adaptation est une façon de susciter l'adhésion des professionnels vis-à-vis d'interventions conçues par des « experts » et vues comme des ingrédients cruciaux d'amélioration, ou si ces interventions sont vues comme des sources parmi d'autres d'inspiration et de réflexivité des professionnels, processus qui seraient *in fine* les vrais ingrédients d'amélioration (McCall, 2009). Le modèle schématique présenté dans la figure 1 et qui sous-tend une bonne part des recherches présentées jusqu'ici, peut en effet être considéré comme reflétant une logique « descendante », dans laquelle des acteurs « externes » disent aux professionnels « ce qu'ils doivent faire » ou quelles sont les « bonnes pratiques ». D'autres modèles ont été proposés pour pallier ce genre de critique. Ainsi, Goigoux et al. (2020) ont proposé d'améliorer ce modèle en associant davantage les professionnels et des préoccupations ergonomiques à la phase de conception, de façon à concevoir des outils didactiques plus robustes. D'autres appellent à la création de partenariats de recherche (Snow, 2015) ou de réseaux de communautés d'amélioration (Bryk, 2015), permettant de développer d'autres relations professionnels-chercheurs. De façon plus radicale, d'autres encore promeuvent un modèle de recherche collaborative, où les professionnels sont impliqués dans toutes les étapes de la démarche (Savoie-Zajc & Descamps-Bednarz, 2007). Ces propositions stimulantes alimentent un débat nécessaire et enrichissent les modèles de recherche disponibles. Notons cependant que ces propositions n'ont pas toutes fait la preuve de leur capacité à entraîner des effets positifs pour les bénéficiaires finaux (souvent les élèves), ni à répondre aux défis des changements à large échelle ou systémiques soulevés plus haut. Comme le soulignent Janosz et Nault-Brière (2020), ces approches demandent non seulement des évolutions dans la formation et les capacités des organisations éducatives, mais aussi des évolutions dans le métier de chercheur et les infrastructures de recherche. Grossman (2009) suggère aussi que, vu la difficulté de développer une action qui ait réellement un effet bénéfique dans la vie des gens, il est peut-être judicieux d'investir davantage dans la dissémination de projets ou d'outils ayant montré leur efficacité potentielle plutôt que dans la réinvention locale de chaque action.

Quoi qu'il en soit, la plupart des chercheurs s'accordent sur le fait que la simple présence de données scientifiques soutenant l'efficacité d'une pratique, ou même la disponibilité de programmes ayant montré leur efficacité, ne garantit en rien leur utilisation par les professionnels. Le déploiement et l'utilisation effective de ces pratiques dans les organisations et dans les communautés, sont liés à d'autres processus, qui ont été beaucoup moins étudiés (Spoth et al., 2013). De plus, l'expérience montre que ces processus demandent du temps et des moyens (Grossman, 2009 ; Noell et al., 2009), d'où l'importance de recueillir des informations sur l'implantation.

De quelle fidélité parle-t-on ?

Au sein de ce débat entre fidélité et adaptation, il est sans doute judicieux de distinguer différentes postures parmi les défenseurs d'une forme de fidélité. Certains chercheurs défendent l'idée d'une conformité à des « programmes exemplaires » ou des « bonnes pratiques », ayant montré leur efficacité dans des études d'évaluation soigneusement réalisées. Ils voient dans l'adoption par les professionnels de ces programmes ou protocoles, labélisés par des organismes ad-hoc, des éléments cruciaux pour améliorer les choses (Slavin, 2002). D'autres chercheurs défendent vigoureusement l'importance de disposer d'études d'évaluation rigoureuses concernant les effets de programmes ou pratiques, ainsi que leur implantation, mais dans l'optique que ces études constituent une base empirique permettant de faire émerger des balises ou principes en vue de soutenir des améliorations, sans conformité rigide à tel ou tel programme ou pratique (Latessa, 2018).

Lipsey (2020) offre une belle synthèse des critiques adressées à l'approche par listes de programmes ou de pratiques « approuvés », « modèles » ou labélisés « fondés sur des preuves » (*evidence-based*). Ces programmes et pratiques ont eu peu d'impact systémique, car ils sont généralement peu adoptés. On peut également s'interroger sur l'adéquation des études qui fondent leur labellisation (études d'efficacité) pour assurer les effets de ces programmes à grande échelle et utilisés dans les pratiques de routine. Les tentatives de réplcation par des chercheurs indépendants des concepteurs sont plutôt rares et produisent souvent des résultats bien inférieurs (études d'impact). Ces pratiques ou programmes « labellisés », parfois réduits aux seuls manuels qui les accompagnent, sont en fait généralement des interventions complexes constituées de multiples éléments. Identifier précisément l'intervention évaluée et la répliquer constitue par conséquent un défi. Très souvent, ces interventions sont considérées comme un « package » avec un seul protocole de traitement, mais dans les faits, des variations importantes sont observées et il est difficile de mettre en évidence les éléments de l'intervention qui sont cruciaux pour produire des effets. Il est fort probable que ces effets puissent dépendre fortement des compétences des professionnels qui délivrent l'intervention et des relations qu'ils établissent avec les bénéficiaires (comme par exemple la notion d'alliance thérapeutique en psychothérapie). Autant d'éléments difficiles à capturer dans un manuel ou un protocole et à répliquer fidèlement en pratique. De plus, il n'existe généralement pas de balises claires pour déterminer à partir de quel degré de variation ou adaptation, on doit considérer qu'il s'agit d'une intervention différente de celle « labellisée ». Les critères de conformité eux-mêmes ont rarement une base empirique qui établit ce qui est critique et à quel degré.

Ces critiques sont cohérentes avec les résultats d'une série convergente de synthèses de recherches qui montrent le faible pouvoir prédictif de la spécificité des interventions. Au terme d'une méta-analyse de projets d'éducation à la résolution de conflit, Garrard et Lipsey (2007) observent un effet moyen positif qui se retrouve, quelles que soient les différentes

approches de résolution de conflit développées dans les projets. Analysant les projets visant à réduire les comportements agressifs et perturbateurs à l'école, Wilson et Lipsey (2007) notent que différentes modalités d'intervention (cognitive, comportementale, fondée sur les compétences sociales, etc.) produisent des effets très similaires, mais que les projets mieux implantés produisent des effets plus prononcés. Les résultats d'une méta-analyse, sur les projets de développement des compétences socioémotionnelles à l'école (Durlak et al., 2011), font apparaître qu'à travers différents types de projets, le respect de certaines balises dans la mise en œuvre du projet et la présence de problèmes d'implantation sont liés aux tailles d'effet observées. Une autre méta-analyse sur les programmes de prévention du décrochage scolaire montre que les effets sont généralement cohérents à travers différents types de programmes, mais qu'un niveau d'implantation plus élevé est associé à des effets plus marqués (Wilson & Tanner-Smith, 2013). Dans une étude sur les programmes de placement judiciaire d'adolescents délinquants, Baglivio et al. (2018) trouvent que la réduction de la récidive est surtout liée à la qualité générale d'implantation plutôt qu'à l'utilisation de tel ou tel traitement spécifique.

Un certain nombre de chercheurs soutiennent donc l'idée que des familles d'interventions similaires ont des caractéristiques communes qui soutiennent l'efficacité sans requérir la conformité à un programme spécifique ou à des pratiques détaillées (Mazzucchelli & Sanders, 2010 ; Wampold & Imel, 2015). Différentes interventions peuvent alors être catégorisées sur la base de similarités fonctionnelles et il devient possible d'étudier systématiquement les variations d'effets entre programmes (notamment au moyen de méta-analyses). Une analyse des programmes de prévention de la récidive montre ainsi qu'il existe une large variance dans la distribution des effets au sein de chaque catégorie de programmes, avec peu de différences entre les programmes « labellisés » et les autres (Lipsey, 2020). Les variables qui sont le plus fortement associées à la taille d'effet de l'intervention, sont les différences de départ entre groupes (niveau de risque), la qualité de la mise en œuvre, la quantité de traitement (dosage), et la modalité de traitement (certains types de programmes sont plus efficaces plutôt que certains programmes spécifiques) (voir aussi Latessa, 2018).

Bref, quand on veut intervenir, les effets semblent dépendre de la quantité d'actions mises en place et de la qualité de celles-ci, bien plus que du choix d'un contenu théorique ou d'une modalité d'action très spécifique (telle procédure précise, tel support, etc.). Cette conclusion rejoint les résultats d'un examen des recherches en santé visant à promouvoir l'activité physique et l'alimentation saine, qui montrent que la théorie n'est que rarement utilisée de façon extensive pour développer les interventions (Prestwich et al., 2014). De plus, ils indiquent que la relation entre le degré d'utilisation de la théorie et l'efficacité est généralement faible. Comme le dit Lipsey (2020), s'appuyer sur un modèle théorique qui n'est pas soumis à des tests empiriques n'est pas en soi un gage de qualité. Rappelons néanmoins que tous les projets analysés dans les études citées ci-dessus, avaient fait l'objet d'un développement soigneux par des personnes expérimentées sur des bases reconnues. Les résultats présentés ne doivent donc pas être interprétés comme soutenant l'idée que tout se vaut en termes d'intervention. Ils suggèrent en revanche qu'au sein d'une gamme d'interventions partageant certaines qualités de base et sous-tendues par une approche commune, l'efficacité dépend en grande partie de la mise en œuvre concrète plutôt que de telle ou telle conduite ou outil spécifique (du moins en lien avec le type de résultats évalués dans ces études, qui n'incluent par exemple pas la maîtrise des matières scolaires).

Pourquoi si peu d'effets observés dans les études d'impact ?

De nombreuses innovations et interventions sont promues par des chercheurs sur la base d'études corrélatives et de modèles conceptuels, en l'absence de données testant leur efficacité potentielle, ou sur la base d'études non répliquées par des équipes indépendantes et/ou réalisées auprès de petits échantillons. D'autres innovations et interventions sont appuyées par des études d'évaluation qui portent sur ce que Prasad et Cifu (2015) appellent des variables de substitution (par exemple, estime de soi, motivation, émotions, stratégies...). Ces variables sont censées affecter le résultat final visé (comportements sociaux, performance, accrochage scolaire, etc.), mais ce résultat n'est lui-même pas mesuré dans ces études. Or, nombre d'études indiquent que promoteurs et usagers peuvent aisément s'illusionner sur l'efficacité d'une intervention (Lilienfeld et al., 2014), que beaucoup d'effets ne se répliquent pas (de Boer et al., 2014 ; De Bruyckere et al., 2019), que les études auprès de grands échantillons montrent des tailles d'effet plus faibles que sur de petits échantillons (Cheung & Slavin, 2016) et qu'une intervention, qui influence une variable de substitution, n'a régulièrement pas d'effet sur l'objectif visé (Melby-Lervåg et al., 2016). Il n'est donc pas tellement étonnant que pareils projets ne produisent pas les résultats escomptés une fois diffusés. En d'autres mots, toute une série d'interventions promues dans les classes et les écoles ne disposent tout simplement pas d'une base suffisamment solide pour justifier l'attente d'effets positifs dans des études d'impact.

Il n'empêche que, comme déjà énoncé précédemment, même des interventions soutenues par des études d'efficacité ont fréquemment des difficultés à montrer des effets observables dans des études d'impact. Comment expliquer cet état des choses ? Certains chercheurs attirent notamment l'attention sur les évolutions qui peuvent apparaître dans les groupes témoins qui servent de comparaison avec les groupes participant à l'intervention (Lemmons et al., 2014). D'autres ajoutent que nombre d'études d'évaluation n'ont pas assez de puissance statistique pour détecter les tailles d'effet qu'on peut habituellement attendre d'une intervention en éducation (Kraft, 2020) ou pour examiner les effets modérateurs de variables de contexte (par exemple, le type d'école) (Spybrook et al., 2020). Au-delà de ces aspects méthodologiques, et comme nous l'avons vu dans ce texte, de nombreux scientifiques soulignent aussi le rôle que jouent les processus d'implantation. Face à la difficulté de retrouver, dans des études d'impact, les effets initiaux trouvés dans des études d'efficacité, plusieurs d'entre eux appellent à concevoir autrement les études d'intervention, de manière à pouvoir dépasser le simple « cela fonctionne » ou « cela ne marche pas » (Jacob et al., 2019). Ils proposent notamment de porter davantage d'attention à documenter les contextes au sein desquels prennent place les interventions et les publics concernés, à mesurer l'implantation de différents éléments de ces interventions et à mesurer également les processus médiateurs supposés expliquer les effets de l'intervention (Kim, 2019). De cette façon, il devient possible d'en apprendre plus sur les processus de changement et de mieux comprendre les raisons qui amènent à observer des effets nuls (Allen-Platt et al., 2021). On peut ainsi examiner les barrières et les adjuvants à l'implantation d'une intervention, l'importance des divers éléments de l'intervention pour les résultats, les caractéristiques des contextes et des publics qui modifient l'impact d'une intervention sur ces résultats, etc. Cette perspective rejoint les deuxième et troisième approches de la mesure de l'implantation présentées plus haut. Cette perspective invite aussi à dépasser le débat entre fidélité et adaptation en considérant que des études d'intervention repensées offrent des perspectives stimulantes pour ouvrir la boîte noire des interventions et tester des théories du changement (Latessa, 2018), en introduisant plus de réciprocité et de boucles de rétroactions dans la figure 1.

Les résultats décevants de beaucoup d'études d'impact constitueraient ainsi une invitation pour les chercheurs à s'intéresser davantage aux organisations et aux professionnels en contact avec les bénéficiaires, aux contraintes et normes professionnelles, aux arrangements organisationnels liés à l'implantation des changements visés, aux critères guidant les jugements professionnels des acteurs concernés, etc. (Goigoux et al., 2021 ; Meyers et al., 2012 ; Ross & Morrison, 2021). Dans cette optique, plusieurs chercheurs invitent à combiner les sciences expérimentales utilisées dans les recherches sur les effets des interventions et les sciences de l'amélioration utilisées dans les études sur l'implantation (Peterson, 2016 ; Spiel et al., 2018).

Conclusion

Au total, il nous semble que deux implications principales ressortent des éléments développés au fil de ce chapitre. Premièrement, vu la complexité du passage d'une avancée dans les connaissances à un éventuel effet positif dans la vie des bénéficiaires, les chercheurs devraient être beaucoup plus prudents dans les implications pratiques de leurs travaux qu'ils mettent en avant. La plupart des recherches ne rencontrent pas les conditions permettant de soutenir de telles implications. Trop souvent, des changements pratiques sont directement promus sur la base d'études transversales ou d'études réalisées « en laboratoire », dont les bénéfices supposés ne sont ensuite pas au rendez-vous, gaspillant ainsi le temps, l'énergie et les ressources des professionnels. Même quand des études d'efficacité existent, une grande prudence reste de mise quant à la capacité des effets observés à être confirmés dans les conditions de routine ou quant à la faisabilité des changements proposés.

Deuxièmement, si on veut mener des recherches ayant davantage d'impact pratique, il est crucial que les chercheurs et les intervenants collectent beaucoup plus systématiquement des informations sur l'implantation des innovations et interventions. L'enjeu central du recueil de ce type d'informations ne serait cependant pas la vérification de la fidélité à un modèle de départ, mais plutôt de se donner l'opportunité de tester les processus théoriques supposés, de fournir des feedbacks aux développeurs en vue d'améliorer les interventions, de construire des modèles de changement plus réalistes par rapport aux contextes réels de mise en œuvre, et de soutenir la durabilité des actions entreprises. Mettre en place à grande échelle des actions qui suscitent un impact positif reste un défi colossal, mais les pistes pour aller de l'avant ne manquent pas.

Références

- Allen-Platt, C., Gerstner, C. C., Boruch, R., & Ruby, A. (2021). Toward a Science of Failure Analysis: A Narrative Review. *Review of Research in Education*, 45(1), 223-252.
- Anderson, E. R. (2017). Accommodating change: Relating fidelity of implementation to program fit in educational reforms. *American Educational Research Journal*, 54(6), 1288-1315.
- Askell-Williams, H., Dix, K. L., Lawson, M. J., & Slee, P. T. (2013). Quality of implementation of a school mental health initiative and changes over time in students' social and emotional competencies. *School Effectiveness and School Improvement*, 24(3), 357-381.
- Baglivio, M. T., Wolff, K. T., Jackowski, K., Chapman, G., Greenwald, M. A., & Gomez, K. (2018). Does treatment quality matter: multilevel examination of the effects of intervention quality on recidivism of adolescents completing long-term juvenile justice residential placement. *Criminology and Public Policy*, 17(1), 147-180.
- Berkel, C., Mauricio, A. M., Schoenfelder, E., & Sandler, I. N. (2011). Putting the pieces together: An integrated model of program implementation. *Prevention Science*, 12(1), 23-33.
- Blakey, C. H., Mayer, J. P., Gottschalk, R. G., Schmitt, N., Davidson, W. S., Roitman, D. B., & Emshoff, J. G. (1987). The Fidelity-Adaptation Debate: Implications for the Implementation of Public Sector Social Programs. *American Journal of Community Psychology*, 15(3), 253-268.
- Bressoux, P. (2020). À quelles conditions peut-on déployer à grande échelle les interventions qui visent à améliorer les pratiques enseignantes ? In B. Galand & M. Janosz (Eds.), *Améliorer les pratiques en éducation : Qu'en dit la recherche ?* (pp. 13-22). Presses universitaires de Louvain.
- Bryk, A. S. (2015). 2014 AERA distinguished lecture. Accelerating how we learn to improve. *Educational Researcher*, 44(9), 467-477.
- Buckley, P., Moore, B., Boardman, A. G., Arya, D. J., & Maul, A. (2017). Validating a fidelity scale to understand intervention effects in classroom-based studies. *American Educational Research Journal*, 54(6), 1378-1413.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation science*, 2(1), 1-9.
- Cheung, A. C., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5), 283-292.
- Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American educational research journal*, 45(2), 443-494.
- Cochrane, W. S., & Laux, J. M. (2008). A survey investigating school psychologists' measurement of treatment integrity in school-based interventions and their beliefs about its importance. *Psychology in the Schools*, 45(6), 499-507.
- Cohen, D. K., Moffitt, S. L., & Goldin, S. (2007). Policy and practice: The dilemma. *American Journal of Education*, 113(4), 515-548.
- Datnow, A., & Castellano, M. (2000). Teachers' responses to Success for All: How beliefs, experiences, and adaptations shape implementation. *American Educational Research Journal*, 37(3), 775-799.
- de Boer, H., Donker, A. S., & van der Werf, M. P. (2014). Effects of the attributes of educational interventions on students' academic performance: A meta-analysis. *Review of Educational Research*, 84(4), 509-545.

- De Bruyckere, P., Kirschner, P. A., & Hulshof, C. (2019). *More urban myths about learning and education: challenging eduquacks, extraordinary claims, and alternative facts*. Routledge.
- De Bruyckere, P., Kirschner, P. A., & Hulshof, C. D. (2015). *Urban myths about learning and education*. Academic Press.
- de Leeuw, R. R., de Boer, A. A., & Minnaert, A. E. (2020). The proof of the intervention is in the implementation; a systematic review about implementation fidelity of classroom-based interventions facilitating social participation of students with social-emotional problems or behavioural difficulties. *International Journal of Educational Research Open*, 1, 100002.
- Dellisse, S. (2020). *Soutenir le changement de pratiques pédagogiques : étude quasi expérimentale sur l'apprentissage de la compréhension à la lecture au début du secondaire*. Thèse de doctorat non publiée. Université catholique de Louvain.
- Dellisse, S., & Galand, B. (2020). L'implantation, un facteur clé dans le changement de pratiques pédagogiques ? In B. Galand & M. Janosz (Eds.), *Améliorer les pratiques en éducation : Qu'en dit la recherche ?* (pp. 23-34). Presses universitaires de Louvain.
- Dellisse, S., Dumay, X., Galand, B., Dupriez, V., Dufays, J. L., Coertjens, L., ... & Penneman, J. (2021). Does the treatment integrity of a literacy instruction program foster student reading comprehension? A field experiment. *European Journal of Psychology of Education*, 36, 573-597.
- Desmet, L., Dupriez, V., & Galand, B. (2017). Que font les écoles accueillant des élèves défavorisés des moyens supplémentaires qui leur sont alloués? Une étude longitudinale de cinq établissements scolaires en encadrement différencié. *Les Cahiers de recherche du Girsef*, (108), 3-91.
- Domitrovich, C. E., & Greenberg, M. T. (2000). The study of implementation: Current findings from effective programs that prevent mental disorders in school-aged children. *Journal of Educational and Psychological Consultation*, 11(2), 193-221.
- Domitrovich, C. E., Bradshaw, C. P., Poduska, J. M., Hoagwood, K., Buckley, J. A., Olin, S., Hunter Romanelli, L., Leaf, P. J., Greenberg, M. T., & Ialongo, N. S. (2008). Maximizing the implementation quality of evidence-based preventive interventions in schools: A conceptual framework. *Advances in school mental health promotion*, 1(3), 6-28.
- Duflo, E. (2020). Field experiments and the practice of policy. *American Economic Review*, 110(7), 1952-1973.
- Durlak, J. A., & DuPre, E. P. (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), 327-350.
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions. *Child development*, 82(1), 405-432.
- Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health education research*, 18(2), 237-256.
- Forehand, R., Dorsey, S., Jones, D. J., Long, N., & McMahon, R. J. (2010). Adherence and flexibility: They can (and do) coexist! *Clinical Psychology: Science and Practice*, 17(3), 258-264.
- Galand, B. (2011). Le rôle de l'environnement scolaire dans le développement des conduites violentes. In F. Bowen & N. Debiens (Eds.), *La violence chez l'enfant : approches cognitive, développementale, neurobiologique et sociale* (pp.199-232). Solal.
- Garrard, W. M., & Lipsey, M. W. (2007). Conflict resolution education and antisocial behavior in US schools: A meta-analysis. *Conflict Resolution Quarterly*, 25(1), 9-38.

- Gentaz, E., Sprenger-Charolles, L., Colé, P., Theurel, A., Gurgand, M., Huron, C., Rocher, T., & 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. *Approche neuropsychologique des apprentissages chez l'enfant (ANAE)*, 123, 172-181.
- Giles, S., Jackson-Newsom, J., Pankratz, M. M., Hansen, W. B., Ringwalt, C. L., & Dusenbury, L. (2008). Measuring quality of delivery in a substance use prevention program. *The journal of primary prevention*, 29(6), 489-501.
- Goense, P. B., Assink, M., Stams, G. J., Boendermaker, L., & Hoeve, M. (2016). Making 'what works' work: A meta-analytic study of the effect of treatment integrity on outcomes of evidence-based interventions for juveniles with antisocial behavior. *Aggression and violent behavior*, 31, 106-115.
- Goigoux, R., Renaud, J. & Roux-Baron, I. (2021). Comment influencer positivement les pratiques pédagogiques de professeurs expérimentés. In B. Galand & M. Janosz (Eds.), *Améliorer les pratiques en éducation : Qu'en dit la recherche ?* (pp. 67-76). Presses universitaires de Louvain.
- Gottfredson, D. C., Cook, T. D., Gardner, F. E., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention science*, 16(7), 893-926.
- Goulet, M., Archambault, I., Janosz, M., & Christenson, S. L. (2018). Evaluating the implementation of Check & Connect in various school settings: Is intervention fidelity necessarily associated with positive outcomes? *Evaluation and Program Planning*, 68, 34-46.
- Gregoire, M. (2003). Is it a challenge or a threat? A dual-process model of teachers' cognition and appraisal processes during conceptual change. *Educational psychology review*, 15(2), 147-179.
- Gregory, A., Henry, D. B., & Schoeny, M. E. (2007). School climate and implementation of a preventive intervention. *American journal of community psychology*, 40(3), 250-260.
- Gresham, F. M., Gansle, K. A., Noell, G. H., & Cohen, S. (1993). Treatment integrity of school-based behavioral intervention studies: 1980-1990. *School Psychology Review*, 22(2), 254-272.
- Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research & Practice*, 15(4), 198-205.
- Grossman, J. B. (2009). Commentary on McCall "Evidence-Based Programming in the Context of Practice and Policy". *Social policy report*, 23(3), 12.
- Haataja, A., Voeten, M., Boulton, A. J., Ahtola, A., Poskiparta, E., & Salmivalli, C. (2014). The KiVa antibullying curriculum and outcome: Does fidelity matter? *Journal of school psychology*, 52(5), 479-493.
- Hamilton, L. S., McCaffrey, D. F., Stecher, B. M., Klein, S. P., Robyn, A., & Bugliari, D. (2003). Studying large-scale reforms of instructional practice: An example from mathematics and science. *Educational evaluation and policy analysis*, 25(1), 1-29.
- Hill, H. C., & Erickson, A. (2019). Using implementation fidelity to aid in interpreting program impacts: A brief review. *Educational Researcher*, 48(9), 590-598.
- Holdsworth, S., & Maynes, N. (2017). "But what if I fail?" a meta-synthetic study of the conditions supporting teacher innovation. *Canadian Journal of Education/Revue canadienne de l'éducation*, 40(4), 665-703.
- Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, 2(1), 88-110.

- Jacob, R. T., Doolittle, F., Kemple, J., & Somers, M. A. (2019). A framework for learning from null results. *Educational Researcher*, 48(9), 580-589.
- Janosz, M., & Nault-Brière, F. (2020). La recherche collaborative avec les milieux de pratique : quelles implications pour le métier de chercheur. In B. Galand & M. Janosz (Eds.), *Améliorer les pratiques en éducation : Qu'en dit la recherche ?* (pp. 91-102). Presses universitaires de Louvain.
- Janosz, M., Bélanger, J., Dagenais, C., Bowen, F., Abrami, P. C., Cartier, S. C., Chouinard, R., Fallu, J. S., Desbiens, N., Roy, G., Pascal, S., Lysenko, L., & Turcotte, L. (2011). Accroître la réussite scolaire en milieu défavorisé : Faits saillants de l'évaluation de la stratégie québécoise d'intervention *Agir autrement. La nouvelle revue de l'adaptation et de la scolarisation*, (4), 203-212.
- Kim, J. S. (2019). Making every study count: Learning from replication failure to improve intervention research. *Educational Researcher*, 48(9), 599-607.
- Klingner, J., Cramer, E., & Harry, B. (2006). Challenges in the implementation of Success for All in four high-need urban schools. *The Elementary School Journal*, 106(4), 333-349.
- Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241-253.
- Latessa, E. J. (2018). Does treatment quality matter: Of course it does, and there is growing evidence to support it. *Criminology and Public Policy*, 17(1), 181-188.
- Lemons, C. J., Fuchs, D., Gilbert, J. K., & Fuchs, L. S. (2014). Evidence-based practices in a changing world: Reconsidering the counterfactual in education research. *Educational Researcher*, 43(5), 242-252.
- Lilienfeld, S. O., Ritschel, L. A., Lynn, S. J., Cautin, R. L., & Latzman, R. D. (2014). Why ineffective psychotherapies appear to work: A taxonomy of causes of spurious therapeutic effectiveness. *Perspectives on Psychological Science*, 9(4), 355-387.
- Lillehoj, C. J., Griffin, K. W., & Spoth, R. (2004). Program provider and observer ratings of school-based preventive intervention implementation: Agreement and relation to youth outcomes. *Health Education & Behavior*, 31(2), 242-257.
- Lipsey, M. W. (2020). Revisited: Effective use of the large body of research on the effectiveness of programs for juvenile offenders and the failure of the model programs approach. *Criminology & Public Policy*, 19(4), 1329-1345.
- Matsumura, L. C., Garnier, H. E., & Resnick, L. B. (2010). Implementing literacy coaching: The role of school social resources. *Educational Evaluation and Policy Analysis*, 32(2), 249-272.
- Mazzucchelli, T. G., & Sanders, M. R. (2010). Facilitating practitioner flexibility within an empirically supported intervention: Lessons from a system of parenting support. *Clinical Psychology: Science and Practice*, 17(3), 238-252.
- McCall, R. B. (2009). Evidence-Based Programming in the Context of Practice and Policy. *Social policy report*, 23(3), 1-20.
- McIntyre, L. L., Gresham, F. M., DiGennaro, F. D., & Reed, D. D. (2007). Treatment integrity of school-based interventions with children in the journal of applied behavior analysis 1991–2005. *Journal of applied behavior analysis*, 40(4), 659-672.
- Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of “far transfer”: Evidence from a meta-analytic review. *Perspectives on Psychological Science*, 11(4), 512–534.
- Meyers, D. C., Durlak, J. A., & Wandersman, A. (2012). The quality implementation framework: A synthesis of critical steps in the implementation process. *American journal of community psychology*, 50(3), 462-480.

- Mihalic, S. F., & Irwin, K. (2003). Blueprints for violence prevention: From research to real-world settings—factors influencing the successful replication of model programs. *Youth violence and juvenile justice*, 1(4), 307-329.
- Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *The American Journal of Evaluation*, 24(3), 315-340.
- Munier, V., Bächtold, M., Cross, D., Chesnais, A., Lepareur, C., Molvinger, K., Gurgand, M., & Tricot, A. (2021). Étude didactique de l'impact d'un dispositif de formation continue à un enseignement des sciences fondé sur l'investigation. *RDST. Recherches en didactique des sciences et des technologies*, (23), 109-136.
- Nault Brière, F. (2017). La recherche d'effectivité : Nature, méthodes et rôle dans la validation des interventions fondées sur les preuves. *Revue de psychoéducation*, 46(1), 117-143.
- Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *The journal of behavioral health services & research*, 39(4), 374-396.
- Noell, G. H., & Gansle, K. A. (2006). Assuring the form has substance: Treatment plan implementation as the foundation of assessing response to intervention. *Assessment for Effective Intervention*, 32(1), 32-39.
- Noell, G. H., & Gansle, K. A. (2009). Moving from good ideas in educational systems change to sustainable program implementation: Coming to terms with some of the realities. *Psychology in the Schools*, 46(1), 79-89.
- Noell, G. H., Witt, J. C., Slider, N. J., Connell, J. E., Gatti, S. L., Williams, K. L., Koenig, J. L., Resetar, J. L., & Duhon, G. J. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review*, 34(1), 87-106.
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of educational research*, 78(1), 33-84.
- Payne, A. A., Gottfredson, D. C., & Gottfredson, G. D. (2006). School predictors of the intensity of implementation of school-based prevention programs: Results from a national study. *Prevention Science*, 7(2), 225-237.
- Peterson, A. 2016. "Getting 'What Works' Working: Building Blocks for the Integration of Experimental and Improvement Science." *International Journal of Research & Method in Education* 39(3), 299-313.
- Powell, B. J., Waltz, T. J., Chinman, M. J., Damschroder, L. J., Smith, J. L., Matthieu, M. M., Proctor, E. K., & Kirchner, J. E. (2015). A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. *Implementation Science*, 10(1), 1-14.
- Prasad, V. K., & Cifu, A. S. (2015). *Ending medical reversal: improving outcomes, saving lives*. Johns Hopkins University Press.
- Prestwich, A., Sniehotta, F. F., Whittington, C., Dombrowski, S. U., Rogers, L., & Michie, S. (2014). Does theory influence the effectiveness of health behavior interventions? Meta-analysis. *Health Psychology*, 33(5), 465-474.
- 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 and mental health services research*, 38, 65-76.
- Quinn, D. M., & Kim, J. S. (2017). Scaffolding fidelity and adaptation in educational program implementation: Experimental evidence from a literacy intervention. *American Educational Research Journal*, 54(6), 1187-1220.

- Rimm-Kaufman, S. E., Larsen, R. A., Baroody, A. E., Curby, T. W., Ko, M., Thomas, J. B., Merritt, E. G., Abry, T., & DeCoster, J. (2014). Efficacy of the responsive classroom approach: Results from a 3-year, longitudinal randomized controlled trial. *American Educational Research Journal*, 51(3), 567-603.
- Ross, S. M., & Morrison, J. R. (2021). Achieving better educational practices through research evidence: A critical analysis and case illustration of benefits and challenges. *ECNU Review of Education*, 4(1), 108-127.
- Rowan, B., & Miller, R. J. (2007). Organizational strategies for promoting instructional change: Implementation dynamics in schools working with comprehensive school reform providers. *American educational research journal*, 44(2), 252-297.
- Rowan, B., Camburn, E., & Correnti, R. (2004). Using teacher logs to measure the enacted curriculum: A study of literacy teaching in third-grade classrooms. *The Elementary School Journal*, 105(1), 75-101.
- Savoie-Zajc, L., & Descamps-Bednarz, N. (2007). Action research and collaborative research: Their specific contributions to professional development. *Educational action research*, 15(4), 577-596.
- Schulte, A. C., Easton, J. E., & Parker, J. (2009). Advances in treatment integrity research: Multidisciplinary perspectives on the conceptualization, measurement, and enhancement of treatment integrity. *School Psychology Review*, 38(4), 460-475.
- Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational researcher*, 31(7), 15-21.
- Snow, C. E. (2015). 2014 Wallace Foundation distinguished lecture. Rigor and realism: Doing educational science in the real world. *Educational Researcher*, 44(9), 460-466.
- Spence, S. H., & Shortt, A. L. (2007). Research Review: Can we justify the widespread dissemination of universal, school-based interventions for the prevention of depression among children and adolescents? *Journal of Child Psychology and Psychiatry*, 48(6), 526-542.
- Spiel, C., Schober, B., & Strohmeier, D. (2018). Implementing intervention research into public policy—the “I3-Approach”. *Prevention Science*, 19(3), 337-346.
- Spoth, R., Rohrbach, L. A., Greenberg, M., Leaf, P., Brown, C. H., Fagan, A., Catalano, R. F., Pentz, M. A., Sloboda, Z., & Hawkins, J. D. (2013). Addressing core challenges for the next generation of type 2 translation research and systems: The translation science to population impact (TSci Impact) framework. *Prevention Science*, 14(4), 319-351.
- Spybrook, J., Zhang, Q., Kelcey, B., & Dong, N. (2020). Learning from cluster randomized trials in education: An assessment of the capacity of studies to determine what works, for whom, and under what conditions. *Educational Evaluation and Policy Analysis*, 42(3), 354-374.
- Tolmatcheff, C., Bebiroglu, N., & Galand, B. (2020). L'aménagement des cours de récréation et l'organisation d'espaces de parole permettent-ils de réduire le harcèlement scolaire ? *Revue française de pédagogie*, (4), 129-147.
- Ttofi, M. M., & Farrington, D. P. (2011). Effectiveness of school-based programs to reduce bullying: A systematic and meta-analytic review. *Journal of experimental criminology*, 7(1), 27-56.
- Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work*. Routledge.
- Webster-Stratton, C., Reinke, W. M., Herman, K. C., & Newcomer, L. L. (2011). The incredible years teacher classroom management training: The methods and principles that support fidelity of training delivery. *School Psychology Review*, 40(4), 509-529.
- West, S. L., & O'Neal, K. K. (2004). Project D.A.R.E. Outcome Effectiveness Revisited. *American Journal of Public Health*, 94(6), 1027-1029.

- Wilson, S. J., & Lipsey, M. W. (2007). School-Based Interventions for Aggressive and Disruptive Behavior: Update of a Meta-Analysis. *American Journal of Preventive Medicine*, 33(2), 130-143.
- Wilson, S. J., & Tanner-Smith, E. E. (2013). Dropout Prevention and Intervention Programs for Improving School Completion Among School-Aged Children and Youth: A Systematic Review. *Journal of the Society for Social Work and Research*, 4(4), 357-372.
- Wilson, S. J., Lipsey, M. W., & Derzon, J. H. (2003). The effects of school-based intervention programs on aggressive behavior: A meta-analysis. *Journal of consulting and clinical psychology*, 71(1), 136-149.
- Wiltsey Stirman, S., Kimberly, J., Cook, N., Calloway, A., Castro, F., & Charns, M. (2012). The sustainability of new programs and innovations: A review of the empirical literature and recommendations for future research. *Implementation science*, 7(1), 1-19.
- Zhang, Y., Fashola, O., Shkolnik, J., & Boyle, A. (2006). Implementation of comprehensive school reform and its impact on increases in student achievement. *Journal of Education for Students Placed at Risk*, 11(3-4), 309-329.