

Des recherches utiles pour la pratique ? Les processus d'implantation comme enjeu transversal en sciences humaines et sociales.

Benoit Galand

Réseau international EVA-3PR, Paris, décembre 2024

benoit.galand@uclouvain.be



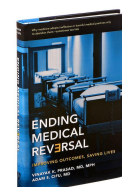
Toutes les questions scientifiquement légitimes n'intéressent pas les praticiens

- La réponse aux questions des scientifiques et celle aux questions des praticiens ont peu à voir l'une avec l'autre (Galand, 2018).
- En sciences, il y a un nombre quasi infini de questions intéressantes.
- Les praticiens (et les bénéficiaires) ont des urgences, des questions qui ont des répercussions importantes dans la « vraie vie ».
- Une question-clé pour le praticien : est-ce que faire quelque chose différemment change le résultat ?
- Souvent des interventions sont promues avant que l'on ait bien étudié leurs effets.

Différents plans/devis de recherche répondent à différentes questions

Étude corrélative transversale	Étude expérimentale	Étude prospective
Revue systématique	Étude par entretiens	Recherche-action
Étude rétrospective	Étude de cas	Méta-analyse
Étude quasi-expérimentale	Études longitudinale	Recherche-développement

Limites des études de cohortes



Nurses' Health Study (NHS)

127 000 infirmières entre 30 et 55 ans ont reçu un questionnaire tous les 2 ans depuis 1976. 90% de taux de réponse.

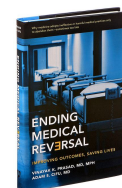
En 1996, publication d'une analyse montrant que celles qui utilisent des oestrogènes ont 40% moins de risque de crise cardiaque que celles qui ne prennent pas d'hormones.

Women's Health Initiative (WHI)

16 000 femmes ayant passé la ménopause (50-79 ans) ont été réparties aléatoirement entre hormones ou placebo. Pairage soigneux concernant l'état de santé. L'étude a été interrompue 3 ans plus tôt que prévu car les femmes qui recevaient des hormones développaient davantage de cancers du sein, de problèmes cardiaques et d'embolies pulmonaires que dans la condition placebo.

NHS – WHI : Une étude observationnelle même bien conçue peut tout à fait nous induire en erreur ! Les femmes qui prenaient des hormones dans l'étude NHS étaient différentes de celles qui n'en prenaient pas...

Le rôle des essais contrôlés randomisés



80'-90', utilisation de substances (ex. flecainide) pour limiter les battements de cœur excessifs (contractions ventriculaires prématurées)

Logique : les CVPs sont fortement associées au risque de décès soudain ; plus vous en avez, plus vous risquez de mourir. Et la flecainide est très efficace pour supprimer les CVPs

En 1992, large essai clinique (CAST) : Répartition aléatoire de façon à ce que les participants soient « identiques » en moyenne. Groupe intervention - Groupe témoin : le plus semblable à l'intervention mais sans l'ingrédient supposé actif (flecainide ou placebo).

Résultats : ces substances réduisent bien les CVPs mais augmentent aussi les risques de décès chez les patients.

Même le meilleur raisonnement et le meilleurs modèles scientifiques ne garantissent pas un traitement clinique efficace.

Si une intervention n'apporte pas de bénéfices, on perd du temps, de l'énergie et de l'argent à la faire durer quand même.

Des interventions adoptées sans preuve d'efficacité nous détournent potentiellement d'alternatives vraiment prometteuses.

Gorman, D. M. (2005). Drug and violence prevention: Rediscovering the critical rational dimension of evaluation research. *Journal of Experimental Criminology*, 1(1), 39-62.

Implicit in these approaches to the identification of science-based prevention programs is the idea that science is equivalent to research methodology and study design. Thus, the better the design of the evaluation, the more scientific it is, and the more confident one can be in the results it produces. In contrast to this, I will argue that, while certain designs are clearly better than others in dealing with threats to internal validity and allow for better generalization of results beyond the study population, **utilization of these designs in and of itself is not sufficient to designate an evaluation study as scientific.** Nor can the accumulation of data from such studies be used to proclaim an entire area of research a science (...)

it is the researcher's willingness to subject his/her theory to falsification and not the use of empirical and observational methods per se that is the key criterion of demarcation between science and other types of knowledge (e.g., common sense, tradition, authority).

Passer d'un résultat à une implication pratique : pas si simple !

- Distinction entre études d'efficacité (*efficacy*) ou de démonstration et études d'impact (*effectiveness*) ou de routine (Gottfredson et al., 2015).
- 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).
- Beaucoup d'interventions, même fondées sur des modèles théoriques bien établis et prometteuses à petite échelle, n'ont pas les effets escomptés une fois diffusée à large échelle (ex. Gentaz et al., 2013 ; Munier et al., 2021 ; Spence & Shortt, 2007 ; West & O'Neal, 2004).
 - Hulleman et Cordray (2009) : 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. La fidélité d'intervention est beaucoup plus faible de la part des enseignants.

7

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

- Beaucoup d'interventions promues dans les milieux professionnels ne disposent pas d'une base suffisamment solide pour justifier l'attente d'effets positifs dans des situations de routine.
 - Pas évaluées
 - Plan de recherche ne permettant pas de tester la direction des effets
 - Petits échantillons
 - Pas répliquées
 - Ciblent des variables « de substitution »
- Rôle que jouent les processus d'implantation dans les études d'impact.

8

TABLE 1
List of Reported Reasons for Null Findings

Events	Reasons Related to Design of Trial		Reasons Related to Design of Intervention	
Planning (59%)	Threats to random assignment	28%	Similar systems already in place	20%
			No single theory of change	10%
			Multiple roles of teachers/schools	14%
			Conflicting goals	10%
			Lack of capacity for change	9%
Implementation (86%)	Intervention period too short	26%	Incomplete implementation	38%
	Study protocol not fully implemented	17%	Dose/intensity of intervention too low	28%
	Follow-up too short	14%	Difficulties/dissatisfaction with resources	28%
	Delays in intervention	10%	Taking time out of teaching content	26%
	Budget constraints	4%	Lack/insufficient training	19%
			Difficult to deliver	16%
			Lack of cooperation/buy-in	12%
			Poor communication/guidance	10%
			Low student engagement	9%
			Low teacher engagement	7%
			Lack of time for cooperation	6%
Instability (41%)	Missing data/low response rate	12%	Attrition of teachers/schools	16%
			Attrition of students	14%
			Low student attendance	14%
			Staffing issues	7%
Measurement (43%)	Issues measuring outcome variables	20%	Allen-Platt, C., Gerstner, C. C., Boruch, R., & Ruby, A. (2021). Toward a Science of Failure Analysis: A Narrative Review. <i>Review of Research in Education</i> , 45(1), 223-252.	
	Contamination of comparison	10%		
	Additional variables missing	9%		
	Sample size too small	9%		
	Sample bias	7%		

Note. The reported reasons for null or negative findings were identified in each study or report, then coded with the four broad categories. Categories were determined a posteriori, as themes emerged from the data. As distinct subcategories emerged, reasons were additionally coded in subcategories. Finally, subcategories were identified by the authors as relating to the design of the trial versus the design of the intervention.

S'intéresser davantage au changements des pratiques professionnelles ?

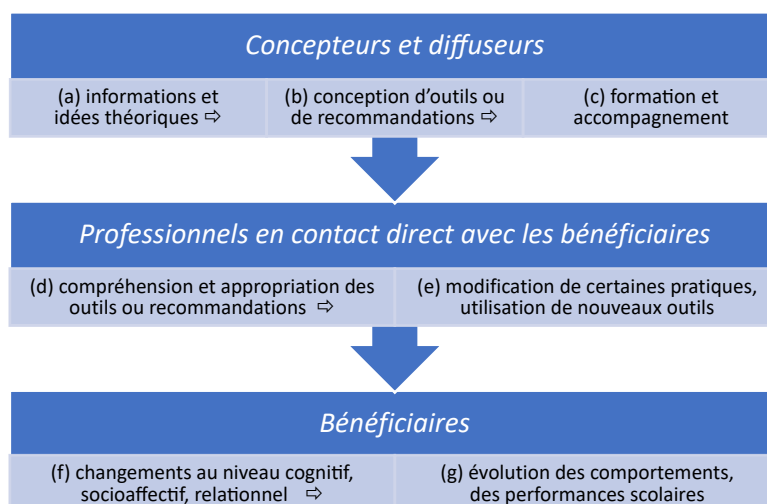
- **Décalage ...**
 - entre connaissances sur les effets de différentes pratiques sur les élèves
 - et connaissances sur les façons de développer ces pratiques chez les professionnels
- La plupart des projets et innovations mise en place par les professionnels n'ont jamais été évalués de façon rigoureuse.
- Dans le même temps, les connaissances issues de la recherche et nombre d'interventions évaluées comme « efficaces » sont très peu utilisées par les professionnels.
- La plupart des réformes échouent à produire les effets attendus (Dupriez, 2015).
- On ignore largement les effets des actions de développement professionnel sur les pratiques et les bénéficiaires (Galand & Dellisse, 2020).
- Quelles recherches pour accompagner des actions à large échelle ayant un impact positif sur leurs bénéficiaires ?

Quel modèle ?

- Modèle « **recherche, développement et diffusion** » (Blakely et al., 1987)
 - 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.
- Trop **simpliste et linéaire** : les acteurs filtrent et transforment activement les outils ou préconisations qui leur sont adressées. Différents acteurs ont des logiques d'action différentes.
- La simple diffusion 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.
- Une constante : les réformes, programmes ou recommandations **ne sont pas mise en œuvre comme planifié**.
- 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).

11

Études translationnelles (Rimm-Kaufman et al., 2014)



12

Trois façons de s'intéresser à la mesure de l'implantation

(Galand et al., 2024)

- A. Mesurer l'implantation pour s'assurer de l'application réelle de l'intervention
- B. Étudier les liens entre degré d'implantation et effets
- C. Identifier et comprendre les facteurs qui influencent l'implantation

13

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

- Mesurer l'implantation = récolter des informations sur ce qui a été fait concrètement par rapport à ce qu'on s'était engagé à faire (Dellisse & Galand, 2020).
- Surtout une question de **validité** (« manipulation check »).
- La majorité des études d'intervention ne mesurent pas la fidélité de la mise en œuvre (Domitrovich & Greenberg, 2000 ; McIntyre et al., 2007), mais cela évolue (Hill et Erikson, 2019).
- La plupart des professionnels 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 n'appliquent pas complètement ou détournent partiellement les démarches ou les innovations proposées (De Smet et al., 2017).

14

B. Étudier les liens entre degré d'implantation et effets

- Une plus grande fidélité d'implantation conduit-elle à des effets plus prononcés (ce qui soutiendrait la pertinence de l'intervention et l'idée que les effets sont causés par l'intervention) ?
- Existe-t-il des seuils de mise en œuvre minimaux ou maximaux afin d'obtenir des effets de l'intervention ?
- Si différentes composantes de l'intervention ont fait l'objet d'une mesure spécifique d'implantation, certaines composantes contribuent-elles davantage aux effets observés ?

15

Association entre implantation et effet

- Relation positive, faible à modérée, entre **degré de fidélité** et **taille des effets**
 - 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 (Askill-Williams et al., 2013), etc.
- Méta-analyse sur plus de 500 études (Durlak & Dupré, 2008) : relation positive entre le **niveau d'implantation** des projets de promotion ou de prévention et les résultats obtenus.
- Les résultats concernant la **durée** des interventions sont nettement moins cohérents.

16

C. Identifier et comprendre les facteurs qui influencent l'implantation

- Enjeu = contribuer à une théorie du changement et dégager des pistes pour améliorer la mise en œuvre d'interventions ou d'innovations.
- Accord sur le fait que des **facteurs à différents niveaux** sont à prendre en compte, mais pas d'accord sur les niveaux et facteurs (Dellisse & Galand, 2020).
 - Dusenbury et al. (2003) : formation des enseignants, caractéristiques du programme, des enseignants et des organisations.
 - Domitrovich et al. (2008) : facteurs au niveau macro, au niveau des écoles et au niveau individuel.
 - Durlak et Dupré (2008) : 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.

17

Facteurs liés au niveau d'implantation

- peu ou pas genre, ancienneté, formation initiale ou composition du groupe d'élèves (Datnow & Castellano, 2000 ; Rowan et al., 2004).
 - croyances, attitudes et expérience préalable des enseignants (Gregoire, 2003 ; Holdsworth & Maynes, 2017).
 - perception par les enseignants que le programme proposé demande des changements de pratiques (Anderson, 2017).
 - standardisation des activités (Rowan & Miller (2007)
 - mise à disposition de matériel détaillé (Hill & Erikson, 2019).
 - collaboration perçue par les enseignants au sein de l'équipe et avec les élèves, leadership administratif de la direction (Gregory et al., 2007).
 - problèmes de regroupement d'élèves et de planning (Klingner et al., 2006).
 - qualité de l'accompagnement, caractéristiques du programme, stabilité des équipes et soutien de la communauté (Mihalic & Irwin, 2003).
- Peu d'études ont directement comparé des facteurs entre eux (Spath et al., 2013).

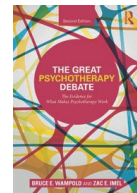
18

Spécificité, fidélité et adaptation

- Deux postures :
- La **conformité** à des « programmes exemplaires » ou des « bonnes pratiques », ayant montré leur efficacité dans des études d'évaluation soigneusement réalisées, et labélisés par des organismes ad-hoc, est cruciale pour améliorer les choses (Slavin, 2002).
- Les études d'évaluation rigoureuses concernant les effets de programmes ou pratiques, ainsi que leur implantation, permettent 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).

19

Relative efficacy



- Statistical theory predicts that by chance some comparisons of treatments will produce statistically significant differences when there are no true differences (i.e. Type I errors).
- This problem is exacerbated by the fact that differences are often found only for a few of the dependent variables in a study.
- (...) there are difficulties in making interpretations from a single comparative study.
- Meta-analysis provides a quantitative test of the hypotheses and avoids conclusions based on salient, but unrepresentative, studies. (p.116-117)

20

Substance abuse



- MATCH (1997): about 1.200 patients randomly assigned to one treatment modality
 - Cognitive-behavioral therapy (CBT)
 - Motivational interviewing (MI)
 - Twelve-steps facilitation (TSF), developed for use in this project and designed to mirror and reinforce the philosophy and strategies offered in AA through individual meetings
- Results: essentially no differences among the treatments
- When directly compared, very different treatments for alcohol-use disorders are equally effective. (p.152)

21

Faible pouvoir prédictif de la spécificité des interventions

- *Education à la résolution de conflit*, effet moyen positif, quelles que soient les différentes approches de résolution de conflit développées dans les projets (Garrard et Lipsey, 2007).
- *Réduction des comportements agressifs et perturbateurs à l'école*, différentes modalités d'intervention (cognitive, comportementale, fondé sur les compétences sociales, etc.) produisent des effets très similaires, mais les projets mieux implantés produisent des effets plus prononcés (Wilson & Lipsey, 2007).
- *Développement des compétences socioémotionnelles à l'école*, 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 (Durlak et al., 2011).
- *Prévention du décrochage scolaire*, les effets sont cohérents à travers différents types de programmes, mais un niveau d'implantation plus élevé est associé à des effets plus marqués (Wilson & Tanner-Smith, 2013).
- *Placement judiciaire d'adolescents délinquants*, 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 (Baglivio et al., 2018).

22

Implications (1)

- 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**.
 - Ex. *prévention de la récidive*, 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. La taille d'effet de l'intervention est principalement associée aux différences de départ entre groupes (niveau de risque), à la qualité de la mise en œuvre, à la quantité de traitement (dosage), et à la modalité de traitement (certains types de programmes sont plus efficaces plutôt que certains programmes spécifiques) (Latessa, 2018; Lipsey, 2020).

23

Implications (2)

- 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).

24

Concevoir autrement les études d'intervention ?

Dépasser le simple « cela fonctionne » ou pas (Jacob et al., 2019), ouvrir la boîte noire des interventions et tester des théories du changement (Latessa, 2018) :

- documenter les **contextes** au sein desquels prennent place les interventions et les publics concernés,
- mesurer l'**implantation** de différents éléments des interventions,
- mesurer les **processus médiateurs** supposés expliquer les effets de l'intervention (Kim, 2019).

- En apprendre plus sur les processus de changement et mieux comprendre les raisons qui amènent à observer des effets nuls (Allen-Platt et al., 2021).
- 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.

25

Conclusions

- Choisir soigneusement ses **questions de recherche** et s'assurer que la méthodologie retenue est bien alignée avec celles-ci.
- **Prudence** quant aux implications pratiques des travaux de recherche.
 - La plupart ne rencontrent pas les conditions permettant de soutenir de telles implications.
 - Les recherches permettent de documenter le réel et pas de décider de ce qui est désirable.
- Pour des recherches ayant davantage de pertinence pratique, collecter beaucoup plus systématiquement des **informations sur la mise en oeuvre**.
L'enjeu n'est pas la vérification de la fidélité à un modèle de départ, mais 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.

26

Pistes de lecture

- Galand, B. (2018). Non, la recherche ne dit pas aux praticiens ce qu'ils doivent faire. *Diversité*, 192, 107-112. https://www.persee.fr/doc/diver_1769-8502_2018_num_192_1_4707
- Galand, B. & Bonnery, S. (2023). « La science c'est la *disputatio*. L'oublier, c'est se fragiliser face aux professionnels comme face aux politiques ». *Diversité* [En ligne], 203 <https://publications-prairial.fr/diversite/index.php?id=4266>
- Galand, B. (2024). « Ce qui me semble beaucoup plus problématique, c'est l'utilisation de résultats de recherche comme argument d'autorité pour donner des injonctions aux professionnels ». *Diversité* [En ligne], 205 <https://publications-prairial.fr/diversite/index.php?id=4552#quotation>
- Meet the lab series: Interview with Dr. Benoit Galand <https://www.baps.be/news/interview/meet-the-lab-series-interview-with-dr-benoit-galand>
- Galand, B., & Dellisse, S. (2021). Comment former les enseignants en cours de carrière pour améliorer les apprentissages des élèves? In Galand, B. & Janosz, M. (dir.), *Améliorer les pratiques en éducation. Qu'en dit la recherche ?* (pp. 57-66). Presses universitaires de Louvain.
- Galand, B., Baudoin, N., & Tolmatcheff, C. (2024). 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.

