

« La boussole de la recherche » : Une introduction à la recherche en éducation médicale : Guide AMEE n° 56

“The research compass”: An introduction to research in medical education: AMEE Guide No. 56

Charlotte RINGSTED¹, Brian HODGES² et Albert SCHERPBIER³

¹ Université de Copenhague et Région Capitale, Danemark

² Université de Toronto, Canada

³ Université de Maastricht, Pays-Bas

Mots-clés

recherche ; éducation médicale ; cadre théorique ; étude exploratoire ; étude expérimentale ; étude d'observation ; étude translationnelle

Résumé – Objectif : Ce guide de l'AMEE est une introduction à la recherche en éducation médicale. Il est destiné aux néophytes qui envisagent de mener une étude dans ce domaine.

Exégèse : Il s'intéresse au processus de transformation des idées et problèmes en questions de recherche, et de choix d'une approche de recherche adaptée aux objectifs de l'étude, qui tient compte tant des préférences du chercheur que des opportunités et contraintes contextuelles. La première partie du guide aborde les raisons de mener une recherche en éducation médicale et les défis posés par la complexité du domaine. La partie suivante explique comment passer d'une idée ou d'un problème à une question de recherche en situant l'idée ou le problème dans un cadre théorique ou conceptuel. Les parties suivantes sont organisées selon « la boussole de recherche », modèle général des approches de recherche en éducation médicale. Au cœur de ce modèle se trouve le cadre conceptuel ou théorique, indispensable pour fixer un cap. La boussole indique les quatre grandes voies d'approches possibles dans l'étude des phénomènes en éducation médicale : les études exploratoires, les études expérimentales, les études d'observation et les études translationnelles. **Perspectives :** Les prochains guides de la série « recherche » aborderont ces approches de façon plus détaillée.

* Cet article est la traduction française d'un travail original élaboré en tant que guide de l'Association for medical education in Europe (AMEE), qui a fait l'objet d'une publication princeps en anglais : Ringsted C, Hodges B & Scherpbier A. "The research compass": An introduction to research in medical education: AMEE Guide No. 56 *Med Teach* 2011; 33: 695–709. Sa publication s'inscrit dans le cadre d'un partenariat institutionnel entre l'Association for medical education in Europe (AMEE) et la Société internationale francophone d'éducation médicale (SIFEM) et leurs organes officiels d'expression respectifs, les revues *Medical Teacher* et *Pédagogie Médicale*. À partir de l'article princeps, la traduction en français a fait l'objet d'une première version, élaborée par Valérie Dory (Institut de recherche santé et société (IRSS), Université catholique de Louvain, Bruxelles, Belgique). La version finale résulte d'un processus de révision collégiale développé ultérieurement au nom du groupe d'action « recherche » de la Société internationale francophone d'éducation médicale (SIFEM) et ayant associé : Valérie Dory, Thierry Pelaccia (Centre d'Enseignement des Soins d'Urgence (CESU), Hospices civils de Strasbourg et Université de Strasbourg, France) et Jean Jouquan (Faculté de médecine, Université de Bretagne occidentale, Brest, France).

Keywords

research; medical education; theoretical framework; explorative study; experimental study; observational study; translational stud

Abstract – Objective: This AMEE Guide offers an introduction to research in medical education. It is intended for those who are contemplating conducting research in medical education but are new to the field. **Exegesis:** The Guide is structured around the process of transforming ideas and problems into researchable questions, choosing a research approach that is appropriate to the purpose of the study and considering the individual researcher's preferences and the contextual possibilities and constraints. The first section of the Guide addresses the rationale for research in medical education and some of the challenges posed by the complexity of the field. Next is a section on how to move from an idea or problem to a research question by placing a concrete idea or problem within a conceptual, theoretical framework. The following sections are structured around an overview model of approaches to medical education research, "The research compass". Core to the model is the conceptual, theoretical framework that is the key to any direction. The compass depicts four main categories of research approaches that can be applied when studying medical education phenomena, "Explorative studies"; "Experimental studies"; "Observational studies"; and "Translational studies". **Further issues:** Future AMEE Guides in the research series will address these approaches in more detail.

Introduction

Ce guide de l'Association for medical education in Europe (AMEE) propose une introduction à la recherche en éducation médicale. Il est destiné à ceux qui envisagent d'entreprendre un travail dans ce champ mais qui y sont, plus ou moins, néophytes. Parmi eux, certains auront peu d'expérience en recherche d'une manière générale tandis que d'autres auront une expérience de recherche dans le domaine biomédical mais pas en éducation médicale. À l'intention de ces derniers, nous établirons des parallèles avec la recherche biomédicale, en en soulignant tant les similitudes que les différences. Ce guide abordera quelques principes généraux et s'intéressera en outre au débat actuel, relatif aux approches alternatives de recherche en éducation médicale et à l'étude des phénomènes et des interventions complexes. Nous espérons ainsi qu'il aura également un intérêt pour les chercheurs en éducation médicale confirmés et qu'il favorisera les discussions futures.

Le guide se centre sur les processus qui permettent de transformer des idées et des problèmes en questions de recherche et de choisir une approche de recherche adaptée aux objectifs de l'étude, qui tienne compte tant des préférences du chercheur que des opportunités et des contraintes contextuelles. La première partie du guide aborde les raisons de mener

une recherche en éducation médicale et les défis posés par la complexité du domaine. La partie suivante explique comment passer d'une idée ou d'un problème à une question de recherche. Elle s'attache en particulier à décrire comment situer une idée ou un problème concrets dans un cadre conceptuel ou théorique. Les parties suivantes sont organisées selon « la boussole de recherche », modèle général des approches de recherche en éducation médicale. Au cœur de ce modèle se trouve le cadre conceptuel ou théorique, indispensable pour fixer un cap. La boussole indique les quatre grandes voies d'approches possibles pour l'étude des phénomènes en éducation médicale : les études exploratoires, les études expérimentales, les études d'observation et les études translationnelles. Le guide comprend des parties dédiées spécifiquement aux principes généraux relatifs, respectivement, à chaque composante du modèle. L'objectif est d'aider les lecteurs dans leur recherche d'informations complémentaires sur leur thématique de prédilection.

Ce guide introductif est le premier d'une série de guides de l'AMEE sur la recherche en éducation médicale. Les guides suivants aborderont des approches de recherche spécifiques, des devis et des méthodes particuliers ; d'autres s'attacheront à des cadres conceptuels ou théoriques et à leur articulation avec l'éducation médicale.

En pratique

- La recherche en éducation médicale cherche à approfondir les connaissances et la compréhension dans le domaine de l'apprentissage, de l'enseignement et de l'éducation. Il ne s'agit ni de résoudre des problèmes concrets locaux ni de fournir des solutions générales et universelles.
- La référence à un cadre théorique ou conceptuel est indispensable pour passer d'une idée, d'un problème ou d'un intérêt à une question de recherche. Il est nécessaire d'identifier les théories rendant compte des mécanismes et des principes d'apprentissage, d'enseignement ou d'éducation associés à la thématique et d'analyser la littérature pour repérer ce que est déjà connu et ce qui doit encore être exploré.
- La recherche implique de faire de petits pas, des choix et des sacrifices, afin de clarifier la thématique étudiée et de formuler une question générale de recherche.
- Il existe de nombreuses approches de recherche visant chacune des objectifs différents. Les quatre catégories principales sont : les études exploratoires qui visent à modéliser ; les études expérimentales qui visent à justifier ; les études d'observation qui visent à prédire et les études translationnelles qui visent à implanter.
- Le choix d'une approche de recherche dépend de la question de recherche. Il est souvent possible et même nécessaire d'en utiliser plus d'une ou de recourir à une approche mixte.

La recherche en éducation médicale

L'objectif altruiste de la recherche en éducation médicale est d'approfondir les connaissances et la compréhension de l'apprentissage et de l'éducation par l'étude des phénomènes et des relations qui permettent de rendre compte de ce qui marche (comment, pourquoi et pour qui ?). On peut toutefois envisager d'autres raisons de s'impliquer dans la recherche en éducation médicale. Le chercheur peut être motivé par

une réelle curiosité intellectuelle quant aux phénomènes en éducation médicale et/ou par un désir de publier à des fins de promotion, de prestige ou de reconnaissance par la communauté scientifique en éducation médicale. Il peut également s'agir de justifier les dépenses, le temps et les efforts investis au profit d'activités en éducation médicale ou, encore, de collecter des fonds destinés à de nouvelles initiatives ou technologies pédagogiques. Enfin, l'engagement dans une approche scientifique en éducation médicale peut être une stratégie pertinente lorsque l'on souhaite convaincre les parties prenantes (*stakeholders*) de nouveaux concepts et idées et obtenir leur soutien aux initiatives innovantes^[1]. Quel que soit le motif de la démarche de recherche, celle-ci nécessite invariablement de la rigueur. Par ailleurs, la publication dans une revue scientifique requiert que l'on identifie clairement l'intérêt pratique de l'étude et que l'on explicite sa portée générale au regard de la connaissance et de la compréhension des déterminants de l'apprentissage et de l'éducation.

Le défi

Le chercheur doit relever le défi de situer une idée, un centre d'intérêt ou un problème concrets dans un contexte plus large d'apprentissage, d'enseignement, et d'éducation. Les chercheurs néophytes ignorent parfois cet objectif de généralisation et ont ainsi tendance à se focaliser sur des problèmes ou des questions concrets d'intérêt local, qu'ils souhaitent comprendre ou résoudre en recueillant et en analysant des données d'évaluations ou d'audits. Par exemple : « Dans quelle mesure les participants à mon cours sont-ils satisfaits de son contenu ou de son format ? » ; « Quel est le résultat d'apprentissage mesuré par l'examen d'habiletés à la fin du cours ? » ou encore « Quel est le degré d'implantation de tel ou tel programme dans mon institution, évalué à travers le nombre d'enseignants adhérant aux standards de référence ? ». Or, l'objectif principal de la recherche n'est pas de répondre à des questions locales concrètes. Au contraire, il s'agit de chercher les réponses à des questions de recherche liées à des questionnements plus généraux sur l'apprentissage,

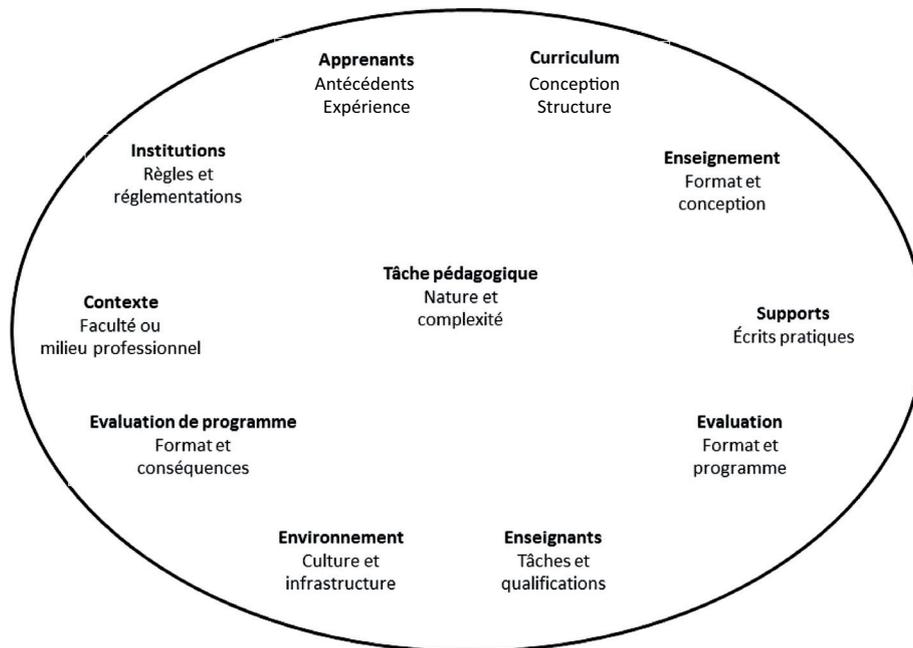


Fig. 1. L'écologie de l'apprentissage. Les facteurs qui interagissent et contribuent à l'apprentissage sont nombreux.

l'enseignement et l'éducation, en étudiant les phénomènes dans des contextes locaux. Dans cette optique, il est primordial de s'intéresser à la généralisabilité des résultats d'une étude et, notamment, d'analyser dans quelle mesure elle apporte une contribution originale en termes de connaissances nouvelles sur l'apprentissage, l'enseignement et l'éducation au sens large. Pour ce faire, il est nécessaire d'évaluer l'impact du contexte de l'étude sur les résultats. Par ailleurs, il est important d'évaluer en quoi ces nouvelles connaissances ont une pertinence pratique et éclairent la pratique pédagogique. Cependant, l'éducation médicale est une discipline très complexe marquée par d'importantes différences de pratiques d'un milieu à un autre, voire au sein même de chaque milieu (cours, écoles, facultés ou autres institutions académiques, sites, juridictions et pays). Cette complexité est encore accrue par les nombreux facteurs qui interagissent pour former l'écologie intriquée de l'éducation (figure 1).

En éducation médicale, la quête d'une solution ou d'une explication qui serait valable dans tous les contextes serait vaine. Une ambition raisonnable pour

la plupart des chercheurs devrait être de s'efforcer d'apporter leur contribution à la construction de nouveaux savoirs, de nouvelles conceptions et de réflexions, visant à comprendre divers phénomènes et à rendre compte des raisons (le « comment » et le « pourquoi ») pour lesquelles certaines choses fonctionnent ou non, ainsi que des conditions nécessaires. En résumé, la recherche en éducation médicale ne consiste ni à résoudre des problèmes strictement locaux, ni à proposer des solutions universelles^[2]. Les chercheurs ont tout intérêt à laisser cela aux parties prenantes et aux praticiens, en espérant qu'ils exploiteront les revues critiques de la littérature scientifique pour améliorer leurs pratiques.

Contrairement aux sciences biomédicales – dont les diverses technologies, qu'elles soient diagnostiques ou thérapeutiques, sont des interventions hautement standardisées –, la discipline « éducation médicale » se caractérise par la non-standardisation de ses pratiques et de ses méthodes. Par ailleurs, les sciences biomédicales bénéficient d'instruments extraordinairement précis pour mesurer les besoins et les résultats. Dès lors, la généralisation est, dans une

certaine mesure, possible et souhaitable. Néanmoins, même dans le domaine biomédical où l'influence d'une conception positiviste du progrès scientifique est majeure, des courants prônent désormais une meilleure prise en compte de la complexité des pratiques et invitent à un élargissement des paradigmes de recherche^[3-5].

De l'idée ou du problème à la question de recherche

Habituellement, c'est la confrontation locale à un problème ou un phénomène concret qui fait naître l'intérêt scientifique chez les chercheurs néophytes dans le domaine de l'éducation médicale. La première tâche du chercheur est alors de transformer ce problème, cette préoccupation ou cette observation en un *problème général de recherche*. Pour ce faire, le chercheur doit identifier et analyser les concepts en jeu dans sa problématique et dans une recherche des mécanismes d'apprentissage, d'enseignement et d'éducation, en rapport avec cette idée. Cette étape est analogue à celle qui consiste, en sciences biomédicales, à cerner les aspects physiologiques, épidémiologiques ou encore, pharmacologiques associés à la pathologie, au médicament ou à la technologie étudiés. En d'autres termes, la première étape de recherche en éducation médicale est de situer l'idée ou le problème dans un cadre théorique ou conceptuel ; en tant que tel, ce processus, par lequel le problème est situé dans son contexte scientifique, correspond à la construction de la problématique de recherche.

Cadre théorique ou conceptuel

Les notions de cadre théorique et de cadre conceptuel sont proches mais pas strictement synonymes, un cadre théorique ayant en général une portée plus large et un caractère plus construit qu'un cadre conceptuel ; un cadre théorique articule d'ailleurs en général plusieurs concepts. Dans la suite du texte, nous parlerons, de façon générique, de cadre théorique ou conceptuel.

Le cadre théorique ou conceptuel d'une étude est constitué de trois éléments : (1) la sélection de

théories de l'apprentissage et de l'éducation permettant de clarifier les mécanismes sous-tendant l'idée ou le problème ; (2) une synthèse critique des données empiriques de la littérature permettant de distinguer ce qui est déjà connu de ce qui ne l'est pas et d'élaborer ainsi une thématique de recherche concrète ; et (3) les réflexions et idées personnelles du chercheur. Ce cadre permet de passer d'une idée ou d'un problème personnel ou local à une problématique de recherche d'intérêt général. Il facilite par ailleurs la formulation d'une question de recherche et le choix d'une approche de recherche adaptée. C'est ainsi que le cadre théorique ou conceptuel structure et organise le processus d'élaboration de la démarche de recherche et de justification des choix méthodologiques à effectuer. Enfin, c'est au regard du cadre théorique ou conceptuel que la discussion scientifique est construite, une fois les résultats obtenus, pour examiner de façon critique leur généralisabilité.

Le défi

L'analyse d'une idée ou d'un problème à la lumière d'un cadre théorique ou conceptuel débute par une étude de la littérature. Pour les néophytes, cette étape est le plus souvent particulièrement exigeante. Elle permet toutefois de réduire *in fine* la charge de travail, en aidant le chercheur à distinguer les voies d'investigation pertinentes de celles qui ne le sont pas. Ce travail commence par l'identification des grands thèmes auxquels renvoient l'idée ou le problème (figure 1). Il s'agit ensuite d'étudier les concepts fondamentaux associés à ces thèmes, ainsi que les mécanismes d'apprentissage, d'enseignement et d'éducation sous-jacents. Ce faisant, le chercheur découvre souvent que son idée ou son problème pourraient être abordés sous différents angles. Des choix doivent donc inévitablement être opérés, le chercheur ayant tout intérêt à choisir la perspective d'étude qui l'enthousiasme le plus.

Où chercher ?

De nombreux ouvrages traitant des concepts fondamentaux de l'apprentissage, de l'enseignement et de l'éducation peuvent être utiles pour identifier le thème

général, l'idée ou le problème à étudier. Il n'existe toutefois aucun ouvrage spécifiquement dédié aux cadres théoriques ou conceptuels, bien qu'il existe plusieurs articles à ce sujet^[5-7]. Par ailleurs, plusieurs articles de synthèse proposent une revue de nombreuses théories en lien avec diverses thématiques en éducation médicale^[8-12]. De nombreux articles de synthèse de la littérature, tels que les revues méthodiques élaborées sous l'égide de la collaboration *Best evidence medical education* (BEME) ou tels que les guides de l'AMEE sont également de bonnes sources de départ pour déterminer ce qui est déjà connu et ce qui ne l'est pas. Ils fourniront de nombreuses références utiles. Ces articles précisent généralement les domaines requérant de nouvelles recherches. Le guide BEME n° 3 (parties 1 et 2) fournit une série de conseils utiles pour effectuer des recherches dans les bases de données pertinentes en éducation médicale^[13, 14]. Le site de l'AMEE (www.amee.org) contient par ailleurs de nombreux liens vers diverses sources de données.

Formuler des questions de recherche

Une fois la thématique de l'étude arrêtée, le chercheur doit formuler une question *générale* de recherche ou un objectif général pour l'étude, qui pourront ensuite être déclinés en questions de recherche plus *spécifiques*. Le format et la formulation des questions sont fortement liés à l'approche de recherche choisie. Les études exploratoires utilisant des méthodes qualitatives posent généralement des questions plutôt ouvertes visant à identifier des variables pouvant expliquer un phénomène, tandis que les études expérimentales posent généralement des questions fermées liées à des variables prédéfinies, visant à documenter une relation. Dans le présent guide, la manière de formuler les questions de recherche sera abordée respectivement dans chacune des sections relatives aux différentes approches de recherche.

La « boussole de la recherche »

La « boussole de la recherche » (figure 2) est un modèle qui propose un aperçu général des approches

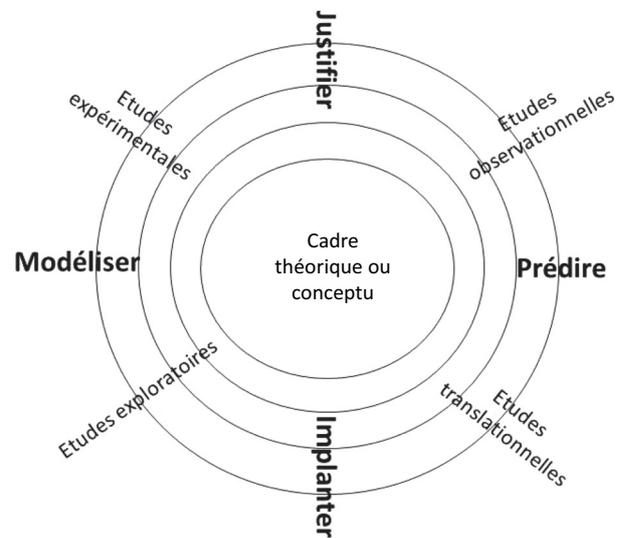


Fig. 2. La « boussole de la recherche ». Un modèle général des approches en matière de recherche en éducation médicale.

utilisées dans le cadre de la recherche en éducation médicale. Ce modèle est issu, d'une part, des perspectives concernant les manières d'aborder les interventions complexes présentées par Campbell et al.^[3] et, d'autre part, de la classification des objectifs de recherche en éducation médicale proposée par Cook et al.^[15] Il est également inspiré de divers ouvrages et articles sur la recherche dans le domaine de l'éducation médicale et des sciences biomédicales.

Le cadre théorique ou conceptuel, qui est le fondement de toute étude et de toute approche de recherche, est représenté au cœur du modèle. Pour aider à s'orienter dans le dédale alentour, la boussole indique quatre grandes catégories d'études, chacune liée à des objectifs différents.

La première catégorie désigne les « études exploratoires » qui visent à modéliser un phénomène éducatif, en cherchant à identifier et expliquer les différents éléments qui permettent d'en rendre compte et à expliciter les relations qui les unissent. Cette catégorie comprend les études descriptives et les études qualitatives, ainsi que les études psychométriques visant à établir la validité et la fidélité des instruments de mesure. La modélisation n'est pas uniquement, en tant que telle, le but des recherches exploratoires, mais peut être aussi l'étape préalable à la conception

d'études expérimentales et à la détermination des interventions et des critères d'évaluation appropriés.

Les « études expérimentales » cherchent à justifier. Ces études sont typiquement des expériences rigoureusement contrôlées avec des groupes homogènes de participants. Ce type d'étude n'est pas toujours réalisable en éducation médicale.

Les « études observationnelles », c'est-à-dire les études de cohorte, les études cas-témoins ou les études visant à établir des associations, représentent souvent une alternative préférable. Les études d'observation examinent des groupes naturels ou statiques de sujets et visent à prédire l'issue.

Enfin, les « études translationnelles » sont centrées sur l'examen de l'implantation du savoir et des découvertes issues de la recherche dans les contextes complexes de la vie réelle, où les sujets ne sont pas similaires entre eux. Cette implantation peut être suivie d'une évaluation tant des processus que des résultats, qui peut elle-même mener à la découverte de nouveaux phénomènes inattendus, conduisant parfois à d'autres recherches exploratoires voire à de nouvelles expériences contrôlées.

Nous ne prétendons pas que le modèle de la « boussole de la recherche » soit exhaustif. Il propose plutôt une vue globale et illustre comment les diverses approches de recherche contribuent, chacune à leur manière, à la construction des savoirs et des conceptions de l'apprentissage et de l'éducation. Contrairement à un labyrinthe, le dédale de la recherche a plusieurs voies d'entrée et plusieurs chemins et issues possibles. Nous ne suggérons pas davantage que tout projet de recherche doive comprendre une approche exhaustive – « à 360° » – d'une thématique. Faire de la recherche, c'est avancer par petits pas, en partant des connaissances et des conceptions existantes. Cela peut se faire à différents endroits du dédale. Le chercheur est libre de se déplacer à tout moment. En d'autres termes, il peut choisir l'approche de recherche qui répond à ses besoins et souhaits, qu'elle soit qualitative ou quantitative, expérimentale ou translationnelle, dès lors qu'elle est appropriée à sa problématique de recherche et réalisable dans le contexte de l'étude. Enfin, les approches de recherche ne sont pas mutuellement exclusives et il n'est pas rare que des combinaisons d'approche soient nécessaires

et réalisables. Nous décrirons plus avant chacun des éléments du modèle dans les sections qui suivent.

Les études exploratoires et la modélisation

Les études exploratoires adoptent des approches de recherche variées, par exemple descriptives, qualitatives et psychométriques. Les études exploratoires ont pour but de modéliser en identifiant, décrivant et analysant les caractéristiques et mécanismes sous-tendant des phénomènes, des comportements, des interventions, des instruments de mesure, etc. Elles posent des questions ouvertes telles que : « qu'est-ce qui caractérise... ? », « comment est-ce que les individus perçoivent ou expliquent... ? », « quels sont les facteurs que l'on peut identifier... ? », « quelle est la validité et la fidélité de... ? ». Sur la base des résultats, le chercheur effectue une synthèse modélisant ou expliquant le thème étudié.

La modélisation désigne également le travail effectué par le chercheur avant de choisir une approche et de concevoir son étude expérimentale, observationnelle ou translationnelle. Dans ce cas, modéliser signifie explorer et analyser des séquences d'une intervention et du mécanisme présumé de son effet dans une étude particulière. La modélisation est également utilisée pour examiner la méthode optimale de recueil des données et de sélection des sujets de l'étude.

Études descriptives

La description de phénomènes ou de nouvelles initiatives concernant, par exemple, l'élaboration de nouveaux programmes, des méthodes d'enseignement, des formats d'évaluation des apprentissages ou des stratégies d'évaluation, n'est généralement pas considérée comme de la recherche, ce qui explique qu'il est de plus en plus difficile de publier ce type d'étude dans des revues scientifiques en éducation médicale. Toutefois, si l'étude aborde une question de recherche adossée à un cadre théorique ou conceptuel, elle sera plus volontiers considérée comme une authentique recherche. Le type de question visée serait, par exemple : « comment les théories de l'apprentissage et de l'enseignement informent-elles l'observation du

phénomène... ? » ou « comment les théories et les études empiriques antérieures informent-elles la conception, la réalisation ou l'évaluation de l'initiative... ? ». Étant donné que la recherche s'intéresse aux « relations », on applique souvent la règle simpliste suivante : pour qu'une étude soit considérée comme ayant le statut de recherche, elle doit concerner une comparaison ou établir des relations. Dans le cas des études descriptives, la « comparaison » peut être développée au regard d'un cadre théorique ou conceptuel.

Par le passé, les études descriptives étaient légion dans les revues scientifiques et lors des conférences en éducation^[16, 17]. Cependant, les leaders de la discipline et les rédacteurs en chef de revues scientifiques ont élevé le niveau de référence des travaux de recherche. Si les études descriptives relatent des constats importants et de bonnes initiatives, y compris des données d'évaluation de programme, d'évaluation des apprentissages ou d'audit, elles manquent souvent d'une perspective de généralisation telle que nous l'avons discutée dans l'introduction de ce guide. Il existe de nombreuses méthodes bien décrites visant à aborder cette perspective de généralisation, telles que l'approche dite de conception itérative « *design-based research* », la recherche-action et l'étude de cas. Une discussion plus approfondie de ces méthodes dépasse le propos de ce guide mais des informations complémentaires sont disponibles dans divers ouvrages ou revues.

Études qualitatives

Les méthodes de recherche qualitatives dans le domaine de la santé et en éducation ont suscité un intérêt croissant au cours des dernières années. Bon nombre de cliniciens formés aux sciences biomédicales sont peu familiers avec les méthodes qualitatives. Selon Pope et Mays^[18], l'objectif général de la recherche qualitative est d'élaborer des concepts permettant de comprendre des phénomènes sociaux dans des contextes naturels (plutôt qu'expérimentaux), en mettant l'accent sur les expériences, les points de vue et le sens donné par les participants. Comme les données qualitatives sont de nature langagière plutôt que numérique, leur lecture, leur interprétation et la

conduite même d'études qualitatives font appel à des méthodes spécifiques^[19].

D'un point de vue théorique, la recherche qualitative est issue d'un courant philosophique apparu au vingtième siècle, qui reconnaît l'importance du langage dans la construction de ce l'on perçoit comme étant la « réalité ». De nombreux chercheurs utilisant les approches qualitatives estiment que ce que l'on appelle « vérité » ou « réalité » est filtré et conditionné par le langage. Dans ce « paradigme constructiviste », les chercheurs utilisant une approche qualitative mettent l'accent sur les contextes sociaux, historiques et personnels^[20].

Les approches qualitatives sont particulièrement indiquées pour répondre à des questions de type « pourquoi », « comment » ou « quelle est la nature de... ». Elles sont utilisées dans trois situations principales : (1) comme préliminaire à une étude quantitative ; (2) en complément de données quantitatives (pour « trianguler » les données) ; et (3) pour explorer des phénomènes complexes qu'une étude quantitative ne pourrait aborder. Il est important de noter que recherche qualitative et recherche quantitative ne s'opposent pas. Bien que certains en fassent des pôles opposés, il s'agit en fait de façons différentes et complémentaires de regarder des phénomènes similaires. Imaginons, par exemple, une étude portant sur le diabète et l'adhérence à la l'insulinothérapie. Une étude quantitative pertinente pourrait explorer la question « quelle est la relation entre la non-adhérence (mesurée par l'hémoglobine A1C) et l'évolution de la maladie ? » Une étude qualitative pourrait explorer les questions « *Pourquoi* les patients n'adhèrent-ils pas au traitement ? *Comment* perçoivent-ils l'évolution de leur maladie ? *Quelle est la nature* de leur conception du diabète et de l'adhérence ? »

Il existe de nombreuses méthodes qualitatives telles que les entretiens, les groupes de discussion focalisée (*focus groups*), les études de cas, l'analyse de texte et l'observation. Il est toutefois important de comprendre les cadres conceptuels associés à chacune de ces méthodes. Chacune appartient à l'une ou l'autre tradition de recherche, telle que la recherche ethnographique^[21], l'analyse de discours^[22] et la théorie ancrée (« *grounded theory* »^[23]). Il est important de bien comprendre ces « méthodologies », dont

découlent les diverses « méthodes » (outils), avant de les utiliser pour recueillir des données.

Bien entendu, les méthodes statistiques ne peuvent pas être utilisées pour examiner des données langagières. Il existe toutefois de nombreuses méthodes pour coder et interpréter les données qualitatives. Généralement, l'analyse débute par le recueil et la transcription des données sous forme textuelle. Kvale a proposé cinq catégories d'approches permettant l'analyse et le codage des transcriptions : la condensation de sens, la catégorisation thématique, la construction de récits, les méthodes interprétatives et enfin une approche composite *ad hoc*^[24].

Une des principales distinctions entre les paradigmes qualitatif et quantitatif est la notion de « qualité »^[25]. Comme les données qualitatives sont recueillies dans des contextes naturels, la priorité étant accordée aux expériences et aux perspectives des participants, aucune présomption de généralisation automatique des résultats à d'autres situations n'est formulée. Les chercheurs évaluent plutôt la possibilité de « transférer » les résultats à d'autres contextes. De même, la notion de « fidélité » des données est remplacée par celle de « crédibilité » et celle de « validité » par celle d'« authenticité ». L'objectif de la recherche qualitative est de comprendre les perspectives et les expériences des participants dans toute leur diversité, plutôt que de réduire les données en éliminant le « bruit » statistique, comme on le fait en recherche quantitative. Les chercheurs parlent de « perspective », incluant la leur – mettant ainsi l'accent sur la « réflexivité » – plutôt que de « biais ». Les autres éléments de qualité concernent le caractère adéquat de l'échantillonnage et la « saturation » des données, l'authenticité et la crédibilité (de bonnes données bien analysées), diverses formes de « triangulation » (examiner les relations et la cohérence de données recueillies de diverses manières et à partir de sources différentes), la « vérification auprès des participants » (restituer les données aux participants pour confirmation), le codage multiple et l'utilisation d'un journal de bord (consignant les décisions et le travail entrepris). Les chercheurs intéressés par les approches qualitatives pourront se référer à un riche éventail de ressources (voir références précitées) et de formations.

Études psychométriques

Les études psychométriques portent d'une manière générale sur la mesure et les instruments de mesure. De nombreuses études concernant l'évaluation ont été publiées au cours des dernières décennies, parmi lesquelles l'immense majorité portait à la fois sur la mesure et les instruments de mesure. Dans cette partie, nous aborderons deux notions fondamentales en matière de mesure, à savoir la validité et la fidélité. Par ailleurs, nous nous attarderons sur l'élaboration d'un « instrument » de mesure ou d'évaluation.

La validité

La validité est une propriété importante d'un instrument de mesure. Un instrument est considéré comme valide lorsqu'il a été établi qu'il mesure bien ce qu'il est supposé mesurer. La validité est bien entendu d'une importance capitale dans les évaluations dont l'enjeu est de prendre des décisions cruciales de réussite ou d'échec. On distingue plusieurs formes de validité dont les quatre principales sont la validité apparente, la validité de contenu, la validité critériée et la validité de construit.

Validité apparente

Par le passé, de nombreuses études ont fait appel à la notion de validité apparente pour indiquer que l'instrument en question semblait à première vue correspondre au contenu qu'il était supposé mesurer. Toutefois, cette définition de la validité n'est généralement plus considérée comme acceptable dans le domaine de la recherche en éducation médicale. Les études qui utilisent cette définition quelque peu naïve de la validité ont peu de chances d'être acceptées par les rédacteurs en chef des revues scientifiques. Ceux-ci exigent en effet des preuves plus tangibles de la validité d'une mesure (voir plus loin). Il y a néanmoins une exception à cette règle : le terme « validité apparente » est en effet couramment utilisé dans la littérature concernant la simulation en éducation médicale pour indiquer que les experts estiment que le modèle présente une approximation acceptable de la réalité de la pratique médicale.

Validité de contenu

L'élément principal de la validité de contenu concerne l'échantillonnage. Ainsi, les concepteurs d'un examen destiné à sanctionner la fin d'un module d'apprentissage devront s'assurer que les questions ou items de l'examen représentent bien le contenu pédagogique du module. Des tables de spécifications sont habituellement utilisées à cette fin. Elles permettent de décrire les différents domaines couverts par le module et de définir le nombre de questions/items par domaine.

La conception d'un questionnaire destiné à mesurer un construit nécessite également que l'on se penche sur la question de sa validité de contenu. Cette conception devrait débiter par une exploration approfondie de la théorie liée au construit et, idéalement, par une recherche d'autres questionnaires déjà publiés. Une fois les thèmes pertinents identifiés et les questions élaborées, il est d'usage de demander à des experts de donner leur avis sur le lien entre le socle théorique et l'instrument, permettant ainsi d'apprécier sa validité de contenu.

Il est regrettable que certains collègues continuent à voir la conception de tests et de questionnaires de manière simpliste. Construire un test et s'assurer de sa validité de contenu est en réalité une tâche exigeante et complexe, notamment en matière de validité de contenu. Dans le domaine de la recherche, il n'est pas rare que les chercheurs soient amenés à élaborer un questionnaire pour recueillir des données sur un sujet peu étudié. Dans ce cas, la meilleure ligne de conduite est de mener en premier lieu une étude qualitative pour mieux cerner l'idée ou le construit et, ainsi, s'informer du contenu. Il est impossible d'évaluer jusqu'à quel point le degré de validité de contenu est atteint. En pratique, cet aspect de la validité est peu abordé dans les publications.

Validité critériée

La validité critériée dépend du degré d'accord entre les résultats du test ou de la mesure en question et ceux d'un autre test ou d'une autre évaluation, désigné comme critère de référence. On distingue deux types de validité critériée : la validité prédictive et la

validité concomitante. Pour évaluer la validité prédictive d'une mesure, on mesure le critère *a posteriori*. C'est le cas, par exemple, lorsque l'objectif est d'évaluer dans quelle mesure les notes du lycée sont prédictives des résultats en faculté de médecine. Nous renvoyons le lecteur intéressé à des exemples bien connus en éducation médicale, tels que les études de Swanson et al.^[26] et de Ramsey et al.^[27].

Lorsque l'on souhaite évaluer la validité concomitante, les deux mesures ont lieu au même moment.

Validité de construit

Il s'agit de l'aspect le plus difficile à apprécier. Pour ce faire, on cherche habituellement à établir des corrélations entre des scores, des mesures et des performances supposées, liées à une théorie ou un construit particuliers. On peut, par exemple, comparer les scores obtenus par les internes et ceux de médecins spécialistes. Comme on suppose que les différences d'expertise auront un effet sur leurs performances respectives, on interprètera une différence de scores entre ces deux groupes comme un argument qui étaye la validité de construit de la mesure. Cette conclusion est en réalité quelque peu simpliste : les différences entre internes et médecins spécialistes sont nombreuses et susceptibles d'influencer leurs performances de diverses manières.

Streiner et Norman^[28] ont décrit trois méthodes fréquemment utilisées pour établir la validité de construit : la comparaison de groupes extrêmes, les études de validité convergente et discriminante et les études multitraits/multiméthodes. Nous renvoyons le lecteur intéressé à leur ouvrage pour en savoir plus.

Fidélité

On distingue deux types de fidélité : la reproductibilité et la cohérence interne. Toutes deux font référence à la constance des résultats de la mesure. Si une mesure est reproductible, cela signifie qu'un étudiant qui obtient un score élevé lors d'un examen obtiendra le même score élevé au même examen présenté à une autre occasion. La cohérence interne implique que les items ou les questions d'un test mesurent toutes la même thématique ou le même construit.

Reproductibilité

La reproductibilité est généralement étudiée par la méthode test-retest ou par la méthode de l'équivalence.

La méthode test-retest repose sur la répétition du même test après un certain délai. La corrélation entre les deux scores est estimée en calculant le coefficient de fidélité qui doit atteindre un seuil minimum. Si l'intervalle entre les deux mesures est trop court, il faut tenir compte de l'éventualité d'un effet d'apprentissage induit par la première mesure. S'il est trop long, de nombreux facteurs seront susceptibles d'influencer les résultats et la corrélation sera plus faible.

La méthode de l'équivalence repose sur l'utilisation simultanée de deux instruments différents pour mesurer les résultats. Les instruments doivent être comparables. Cette méthode est moins souvent utilisée que la méthode test-retest, en raison de la difficulté à élaborer deux instruments réellement comparables entre eux.

Dans la littérature sur l'évaluation, on distingue deux types de reproductibilité : la fidélité intrajuges et la fidélité interjuges. Lorsque le candidat est évalué par un ou plusieurs juges, la question de la fidélité intrajuge peut se formuler comme suit : « les jugements émis par le même évaluateur sont-ils constants dans le temps ? ». En d'autres termes, l'évaluateur portera-t-il le même jugement sur le candidat après un certain laps de temps (fidélité intrajuge) ? La question qui se pose pour la fidélité inter-juge est la suivante : « les différents évaluateurs émettent-ils le même jugement concernant un même candidat ? ».

Cohérence interne

La cohérence interne d'un test ou d'une mesure dépend du degré d'homogénéité entre les items d'une même échelle. En d'autres termes, il s'agit de savoir dans quelle mesure les items mesurent le même construit ou concept sous-jacent. Il existe deux manières d'évaluer la cohérence interne : calculer le coefficient de fidélité ou diviser le test en deux et calculer le coefficient de chaque partie.

Le calcul du coefficient alpha de Cronbach est le test statistique le plus fréquemment utilisé pour évaluer la cohérence interne. Le coefficient de Kuder-Richardson est plus adapté en cas de variables

dichotomiques. Ces coefficients ne sont pas des propriétés stables d'un instrument donné. Ils sont liés à une mesure donnée à un moment donné et sont donc susceptibles d'évoluer. Une autre manière de procéder est de diviser le test en deux et de vérifier si les deux moitiés produisent les mêmes résultats. La formule de Spearman-Brown peut être utilisée à cet effet.

L'élaboration d'un instrument

Le processus d'élaboration d'un instrument commence par une description et une définition claires de la thématique étudiée, de ce que l'on cherche à mesurer et de la manière dont on envisage de le faire. Une analyse de la littérature s'impose. Elle a plusieurs objectifs : 1) la recherche de modèles théoriques ou de théories en lien avec la thématique, permettant d'accréditer l'instrument ; 2) l'identification de ce que d'autres ont étudié en lien avec la thématique et 3) la recherche d'instruments existants et la détermination de leurs validité et fidélité.

Le chercheur novice pourrait utilement discuter de ses projets d'élaboration d'un instrument avec des collègues plus expérimentés familiers de l'état des connaissances et susceptibles de lui prodiguer des conseils quant aux articles importants à lire. Les chercheurs expérimentés déconseilleront souvent aux novices de créer un instrument *de novo*.

Une fois la thématique bien décrite et une définition claire formulée, le chercheur ayant décidé d'élaborer un nouvel instrument s'attachera à définir le contenu à mesurer. Il existe différentes manières de définir un contenu et d'élaborer une table de spécification en fonction de la thématique. Par exemple, si l'instrument vise à fournir des informations sur une procédure, l'analyse de celle-ci pourrait servir de point de départ. L'état des connaissances n'est pas toujours suffisant pour définir le contenu de la mesure. Dans ce cas, une étude qualitative préliminaire utilisant, par exemple, des entretiens ou des groupes de discussion focalisée est indiquée.

On utilise souvent les réunions d'experts pour définir le contenu de la mesure. Les fondements théoriques sont importants à prendre en compte à cette étape de la conception d'un instrument. La théorie

intervient dans les décisions au sujet du contenu, en guidant les choix concernant l'angle ou la perspective que le chercheur adoptera pour aborder la thématique.

Une fois le contenu défini, le chercheur peut commencer à construire son instrument. Il pourrait s'avérer précieux de le soumettre à des collègues pour vérifier les formulations, l'absence de redondances, etc.

Une fois la première version de l'instrument produite, celui-ci peut être mis à l'épreuve. Il est important de déterminer si les « utilisateurs » comprennent bien les questions et combien de temps il leur faut pour y répondre. Dans les grands projets de recherche, cette phase pilote concernera un nombre suffisant de participants pour permettre des analyses statistiques. Une fois l'instrument amélioré, suite aux résultats de la phase pilote, il est prêt à être administré (encadré 1).

Pour déterminer quels tests seront appliqués aux données recueillies, il faut déterminer le type d'échelle concerné : s'agit-il d'une échelle nominale, ordinale, à intervalles ou proportionnelle ?

Une échelle nominale est une échelle où un chiffre est attribué aux variables par le chercheur. Par exemple : 1 = homme et 2 = femme. Ces chiffres n'indiquent ni une valeur numérique, ni un ordre particulier. Ils servent uniquement à placer les données dans une catégorie. Dans une échelle ordinale, l'information est classée dans un certain ordre mais les différences entre les points de l'échelle ne sont pas pour autant égales. Dans une échelle à intervalles, les distances entre les points de l'échelle sont toutes égales. Par exemple, la distance entre 10 et 20 kilos est la même qu'entre 40 et 50 kilos. Il existe de nombreuses circonstances où l'on ne peut pas affirmer avec certitude que les intervalles entre les points de l'échelle sont égaux ou non. Une échelle proportionnelle est une échelle d'intervalles avec une vraie valeur zéro, comme pour la taille ou le poids. En pratique, les chercheurs supposent souvent que leur échelle est à intervalles parce que ce type d'échelle se prête à des analyses statistiques plus variées.

Les études expérimentales

L'objet des études expérimentales est la justification. En d'autres termes, elles visent à mettre en évidence les effets d'une intervention. Les questions de

Encadré 1. Élaboration d'un instrument.

1. Analyse de la littérature
2. Définition d'une thématique d'étude
3. Décision concernant le contenu
4. Construction de l'instrument
5. Pilotage et amélioration de l'instrument
6. Administration de l'instrument

recherche sont typiquement fermées : « Est-ce que cette intervention fonctionne ? » ou « Est-ce que l'intervention A est plus efficace ou plus efficiente que l'intervention B ? ». Pour s'assurer que l'effet observé est bien dû à l'intervention en question, le chercheur doit contrôler toute une série de facteurs susceptibles d'influencer les résultats. Ces facteurs, dits confondants, peuvent concerner les caractéristiques de la population étudiée, l'intervention elle-même ou encore, le contexte de l'étude. En contrôlant rigoureusement son expérimentation, le chercheur s'efforce de minimiser les effets indésirables des facteurs confondants. Il existe plusieurs manières de procéder : sélectionner un échantillon homogène, standardiser l'intervention et l'instrument de mesure, prévoir un groupe contrôle non soumis à l'intervention. Dans les exemples suivants de devis expérimentaux, nous aborderons de manière générale les difficultés et les défis auxquels la recherche expérimentale en éducation médicale doit faire face. Nous nous limiterons toutefois à exposer des principes généraux et renvoyons le lecteur à la littérature concernant les aspects méthodologiques précis relatifs à la conduite d'une étude expérimentale, notamment en matière d'analyses statistiques.

Il existe de nombreux types d'études expérimentales. Dans le domaine biomédical, l'essai randomisé contrôlé (*randomised controlled trial* – RCT) est généralement considéré comme le devis souverain, permettant mieux que tout autre d'obtenir des preuves valides de l'effet d'une intervention. Le RCT permet d'étudier si une intervention fonctionne dans des situations strictement contrôlées et standardisées. L'utilisation de critères d'inclusion et d'exclusion stricts, ainsi que la répartition aléatoire des sujets dans le

Tableau I. L'essai randomisé contrôlé (*Randomised controlled trial* – RCT).

RCT classique				
Randomisation	Groupe I Groupe C	M avant M avant	I C Placebo	M après M après

Note : I = intervention, C = contrôle et M = mesure.

Tableau II. Devis d'un essai contrôlé randomisé à quatre groupes avec mesures ultérieures.

Devis de Solomon	Suivi					
Randomisation	I Groupe 1 C Groupe 1 I Groupe 2 C Groupe 2	M avant M avant	I nouvelle I alternative I nouvelle I alternative	M après M après M après M après	M rétention M rétention M rétention M rétention	M transfert M transfert M transfert M transfert

Note : I = intervention, C = contrôle et M = mesure.

groupe d'intervention ou le groupe de contrôle sont supposés garantir que les facteurs confondants liés à des caractéristiques individuelles seront identiques dans les deux groupes et que la différence observée réside uniquement dans le fait d'être soumis ou non à l'intervention. Le RCT classique compare intervention et absence d'intervention (placebo) (table I). Les participants sont répartis de manière aléatoire dans le groupe intervention ou le groupe contrôle. Une mesure initiale est effectuée dans les deux groupes, suivie de l'intervention dans un groupe et d'une intervention placebo dans l'autre et, enfin, d'une mesure finale dans les deux groupes. Les données sont ensuite analysées à la recherche de différences au sein des groupes et entre les groupes, l'hypothèse étant qu'un changement aura eu lieu dans le groupe intervention mais pas dans le groupe contrôle.

Le devis classique de RCT pose trois problèmes dans le domaine de l'éducation médicale. Tout d'abord, cela n'a fondamentalement pas de sens de comparer une intervention à une absence d'intervention en éducation. Si l'intervention concerne, par exemple, une formation à la réanimation, le groupe soumis à l'intervention ne peut que faire mieux que le groupe contrôle qui n'a pas suivi le cours puisque l'apprentissage de la réanimation ne peut résulter d'une évolution spontanée. En éducation médicale, la comparaison d'une nouvelle intervention à une

alternative raisonnable plutôt qu'à l'absence d'intervention a davantage de sens.

Le deuxième problème soulevé par le devis schématisé dans le tableau I est l'influence potentielle de la mesure initiale sur le résultat, *via* un apprentissage par le test (*test-enhanced learning*), *via* la motivation à apprendre ou encore, *via* l'activation de connaissances antérieures. Par ailleurs, si le même test est utilisé avant et après l'intervention, il est possible que les participants s'en souviennent. Il faut donc être prudent lorsqu'on utilise des pré-tests dans le domaine de l'éducation médicale. Le chercheur est toutefois confronté à un dilemme : comment est-il possible de quantifier les apprentissages sans pré-test ?

Enfin, le temps écoulé entre l'intervention et la mesure finale peut influencer considérablement les résultats. Si l'apprentissage est mesuré immédiatement après l'intervention, les résultats peuvent être trompeurs. Un réel apprentissage suppose un changement de capacité relativement durable, qu'il s'agisse de connaissances, d'habiletés, d'attitudes, de motivation, d'émotions ou de socialisation. Pour déterminer qu'un apprentissage durable a eu lieu, il est recommandé de procéder à des mesures de suivi en évaluant la rétention des apprentissages à distance de l'intervention et/ou la capacité à appliquer leurs acquis dans d'autres tâches ou d'autres contextes, c'est-à-dire, le transfert des apprentissages.

Le devis RCT à quatre groupes permet de prendre en compte ces trois difficultés (tableau II). Les participants sont répartis en quatre groupes. Seuls les deux premiers groupes sont soumis à une mesure initiale. Le tableau indique par ailleurs que l'intervention est comparée à une alternative appropriée et qu'une mesure de suivi est prévue.

Le devis RCT à quatre groupes représente un idéal qu'il n'est pas aisé d'atteindre en pratique. Une alternative réaliste est de diviser l'étude en une étude pilote, prévoyant une mesure avant et après l'intervention, et une étude principale sans mesure initiale mais avec des mesures de suivi. Enfin, la randomisation peut s'avérer difficile en éducation médicale. Le chercheur devra souvent se contenter d'autres méthodes de composition des groupes à comparer. Ces études non-randomisées sont également appelées études quasi-expérimentales.

Le défi

Quel que soit le devis, RCT ou étude quasi-expérimentale, les études expérimentales sont confrontées à des défis qu'il s'agit de relever avant de pouvoir formuler une question de recherche spécifique. Ces défis relèvent de six éléments imbriqués : le contexte, la population étudiée, les conditions de l'intervention et du contrôle, la mesure du résultat et le temps de mesure. Ces éléments, identifiables par l'acronyme SPICOT (pour *Setting, Population, Intervention, Control, Outcome, Time*)^[29], doivent tous figurer dans la question de recherche spécifique d'une étude expérimentale. Bien que l'on admette couramment que c'est la question de recherche qui détermine l'étude, l'identification et la définition des éléments SPICOT doivent être comprises comme un processus itératif de modélisation cohérente avec le cadre conceptuel ou théorique choisi et tenant compte de l'imbrication des éléments SPICOT. La modification d'un élément a toujours des répercussions sur au moins l'un des autres éléments. Le choix des éléments SPICOT dépendra également des contraintes pratiques liées au contexte de l'étude. Des arbitrages, voire des sacrifices, devront inévitablement être opérés. La perfection n'existe malheureusement pas en matière d'études expérimentales.

Toutefois, de la rigueur exercée pour clarifier et décrire chacun des éléments ainsi que des choix effectués dépendront la qualité de l'étude et la possibilité d'en tirer des conclusions et des publications.

Comment définir les éléments SPICOT

Il s'agit avant toute chose de définir soigneusement les conditions de l'intervention et de l'intervention alternative ou du contrôle. Cette tâche s'avère particulièrement délicate en éducation médicale où une intervention, une formation ou un cours comprennent généralement plusieurs composantes ayant chacune leur impact sur l'apprentissage. De plus, il n'existe pas de terminologie consacrée en éducation médicale. Les concepts d'« apprentissage par problèmes » ou d'« enseignement didactique traditionnel » peuvent recouvrir des réalités bien différentes selon le contexte. Il est dès lors essentiel de bien décrire les interventions prévues. Il peut aussi être utile d'émettre des hypothèses concernant les raisons et les modalités de l'impact potentiel de l'intervention, ainsi que les raisons d'un éventuel échec. Ces questions devraient être abordées lors de la phase de formulation de la problématique, avant la conduite de l'étude proprement dite.

Une fois les conditions d'intervention et de contrôle définies, l'étape suivante consiste à identifier ou élaborer une mesure valide du résultat et à définir le temps de mesure. Les défis associés à cette étape ont été abordés dans les parties sur les études psychométriques et les difficultés liées aux RCT.

Enfin, l'échantillon devra être déterminé de manière appropriée, tant sur le plan quantitatif que qualitatif. Concernant la quantité, le nombre de participants dépend de la taille de l'effet attendu, de la sensibilité de l'instrument de mesure et de la variance parmi les participants. Des calculs de puissance peuvent être utilisés pour estimer la taille d'échantillon requise. Le lecteur est renvoyé à des ouvrages de statistique pour de plus amples informations sur le sujet mais, de manière générale, plus la taille de l'effet attendu est grande, plus l'instrument est sensible et moins il y a de variance dans l'échantillon, plus la taille d'échantillon requise pour conclure est petite. Dans les études expérimentales, il est courant

de définir des critères d'inclusion et d'exclusion stricts afin d'obtenir un échantillon le plus homogène possible. La similarité au sein de l'échantillon peut toutefois limiter la généralisation des résultats à d'autres collectifs, tels qu'ils se présenteront en contexte authentique. En d'autres termes, cela peut nuire à la validité externe de l'étude. L'exemple suivant illustre ce phénomène. Les études expérimentales concernant les simulateurs haute-fidélité sont souvent menées sur des échantillons d'étudiants en médecine. Ces échantillons ont l'avantage d'être assez homogènes quant aux apprentissages antérieurs puisqu'il s'agit de novices par rapport à la procédure visée par l'entraînement sur simulateur. Toutefois, les principes pédagogiques dérivés d'une étude menée chez des novices pourraient être différents, voire contradictoires à ceux qui devraient être appliqués avec des apprenants plus avancés ou plus expérimentés^[29, 30]. Ce risque existe tant pour les aspects cognitifs et émotionnels de l'apprentissage, que pour ses aspects sociaux^[31]. L'instrument de mesure peut également poser problème, car il peut être approprié pour mesurer les phases précoces de l'apprentissage mais pas nécessairement les phases plus avancées.

Les études observationnelles

La prédiction d'une relation entre variables est l'objet central des études observationnelles. Ces études diffèrent des études expérimentales par plusieurs aspects. Les participants sont des groupes naturels ou statiques plutôt que des échantillons sélectionnés, même si des critères d'inclusion et d'exclusion sont généralement également appliqués. Dans les études observationnelles, les participants ne sont pas répartis dans un groupe par le chercheur : les groupes sont définis, par exemple, par les promotions d'une faculté ou par l'année d'inscription.

Il y a plusieurs raisons de choisir ou non une étude d'observation plutôt qu'une étude expérimentale. Il peut d'abord s'agir de raisons liées à la faisabilité d'un RCT : difficultés à soumettre un groupe expérimental et un groupe contrôle à des interventions différentes, résistance de la part des étudiants/autorités qui auraient des convictions personnelles concernant les

avantages/inconvénients d'une des interventions proposées, difficultés de recrutement des participants dans un délai raisonnable. À titre d'exemple, le recrutement peut s'avérer problématique dans les cursus médicaux post-gradués, vu le nombre restreint d'internes dans chaque spécialité, par rapport au nombre d'étudiants de première année. Par ailleurs, si les études expérimentales sont très utiles pour démontrer qu'une intervention fonctionne dans un groupe de personnes rigoureusement sélectionnées et dans des circonstances strictement contrôlées, les RCT souffrent généralement d'une piètre validité externe, c'est-à-dire que leurs résultats ne permettent pas de prédire que l'intervention fonctionnera effectivement dans des groupes réels moins homogènes. Enfin, les études d'observation sont plus adaptées à des questions de recherche s'inscrivant dans une perspective à long terme et tenant également compte d'effets involontaires.

Les types d'études observationnelles

Ce guide abordera brièvement trois types d'études observationnelles : les études de cohorte, les études cas-témoins et les études d'association. La différence entre les études de cohorte et les études cas-témoins concerne la direction de l'investigation (figure 3). Dans les études de cohorte, l'investigation commence par la variable indépendante (prédicteur), c'est-à-dire une exposition ou un facteur qui différencie les deux cohortes, tandis que dans les études cas-témoins, l'investigation commence par la variable dépendante (critère), c'est-à-dire une caractéristique résultante différente entre les groupes. Le chercheur est satisfait si la majorité des sujets de l'étude conforte l'hypothèse d'une relation entre les variables dépendante et indépendante (situation représentée par les quadrants supérieur gauche et inférieur droit de la figure). Le chercheur est déçu si les sujets tombent dans les quadrants supérieur droit ou inférieur gauche, qui sont en contradiction avec l'hypothèse d'une relation entre variables dépendante et indépendante.

Les études de cohorte

Dans les études de cohorte, l'investigation commence par la variable indépendante (prédicteur), c'est-à-dire

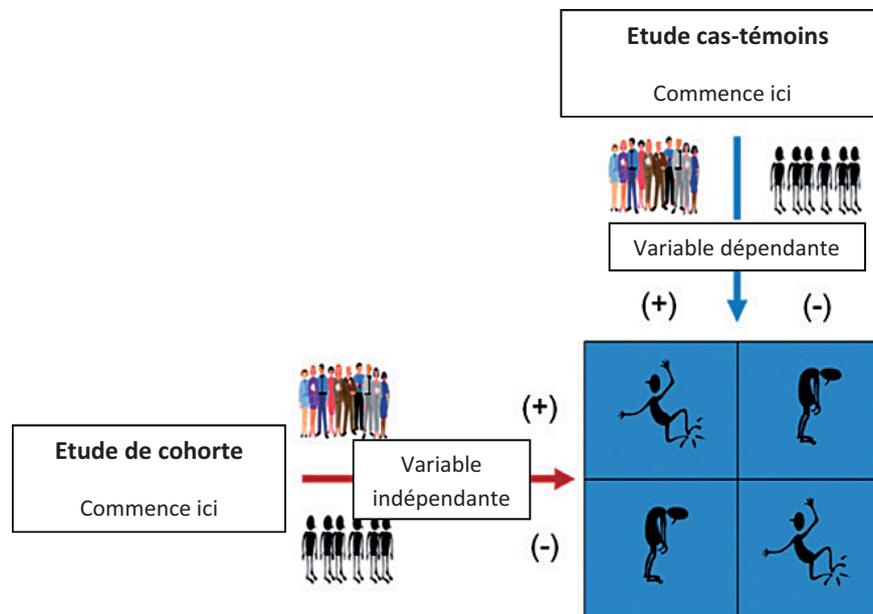


Fig. 3. La différence entre les études de cohorte et les études cas-témoins concerne la direction de l'investigation.

par l'exposition de la cohorte à une activité pédagogique ou les caractéristiques spécifiques de la cohorte, et explore la relation entre cette variable indépendante et la variable dépendante (le critère ou la caractéristique résultante). La question de recherche pourrait être « est-ce que le fait d'avoir suivi des études dans un curriculum fondé sur l'apprentissage par problèmes (APP) est prédictif de la compétence médicale future ? »^[32]. La variable indépendante est l'exposition à l'APP et la variable dépendante est la compétence comme praticien, mesurée par des pairs. Des cohortes d'étudiants de facultés non-APP sont comparées à des cohortes d'étudiants de facultés APP. C'est la variable indépendante qui définit les groupes. Notons au passage qu'une étude de cohorte rétrospective portant sur cette question n'a pas mis en évidence de relation de ce type^[32].

Les études de cohorte peuvent également suivre une cohorte dans le temps et contrôler les résultats sur des variables dépendantes prédéfinies. Par exemple, les résultats d'un examen en fin d'études de médecine (ici, la variable indépendante est une caractéristique plutôt qu'une exposition) pourraient être étudiés quant à une éventuelle association avec des données

concernant la qualité des soins (variables dépendantes) prodigués par la même cohorte, une fois que celle-ci sera en situation de pratique^[33, 34].

Les études cas-témoins

Dans les études cas-témoins, l'*investigation* commence par la variable dépendante (critère) et examine l'exposition ou les caractéristiques des participants associées à cette caractéristique résultante (les cas), par rapport à un groupe de participants appropriés (les témoins). C'est donc la variable dépendante qui définit les groupes. À titre d'exemple, la variable dépendante pourrait être les mesures disciplinaires prises par l'ordre des médecins et la variable indépendante les notes concernant le professionnalisme, obtenues pendant les études^[35]. Dans cette étude, le groupe de cas comprenait des médecins ayant fait l'objet de mesures disciplinaires et le groupe de témoins incluait des médecins n'ayant pas fait l'objet de telles mesures. Les groupes étaient comparables sur le plan de la faculté d'origine, de l'année d'obtention du diplôme et du choix de spécialisation. Les données relatives au professionnalisme de chaque groupe

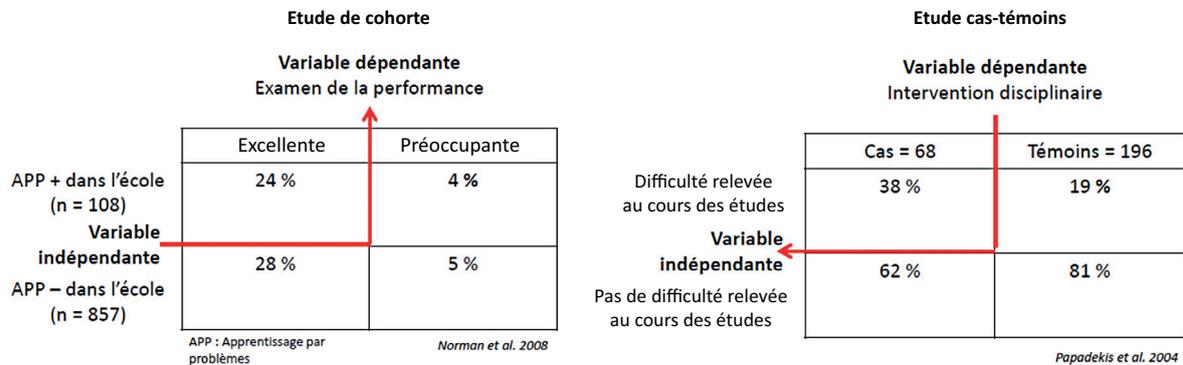


Fig. 4. Exemples d'une étude de cohorte et d'une étude cas-témoins.

durant les études ont été analysées et comparées. En l'occurrence, cette étude a mis en évidence l'existence d'une relation^[35].

La figure 4 schématise l'étude de cohorte de Norman et al.^[32] et l'étude cas-témoins de Papadakis et al.^[35]. L'étude de cohorte n'a pas mis en évidence de relation entre la formation par l'APP et l'évaluation ultérieure comme « excellente » ou « préoccupante ». Les proportions de sujets dont la compétence était considérée comme « excellente » et comme « préoccupante » étaient les mêmes dans les deux cohortes. L'étude cas-témoins a quant à elle mis en évidence une prévalence de la variable indépendante « craintes concernant le comportement professionnel relevées pendant les études » deux fois plus élevée dans le groupe de cas que dans le groupe de témoins, indiquant une association entre la variable indépendante et la variable dépendante, à savoir, les mesures disciplinaires prises par l'ordre des médecins.

On peut plus rapidement mener une étude cas-témoins ou une étude de cohorte rétrospective qu'une étude prospective. Les études cas-témoins sont particulièrement utiles lorsque la variable dépendante est nominale (oui/non, présente/absente), lorsque la caractéristique résultante est rare et son délai d'apparition long. On considère toutefois que la robustesse des preuves issues d'une étude cas-témoins est inférieure à celle de preuves issues d'une étude prospective en ce qui concerne une éventuelle relation causale. Une étude cas-témoins peut néanmoins être un bon point de départ pour explorer l'association entre une série de variables dépendantes et indépendantes,

association qui pourra ultérieurement faire l'objet d'études plus robustes.

Les études d'association

Les études d'association ne comparent pas nécessairement des groupes. On peut citer, par exemple, les études transversales qui livrent une photographie de certaines variables chez diverses personnes et examinent leurs associations (figure 5).

La variable indépendante pourrait être l'expérience clinique mesurée en années de pratique et la variable dépendante la qualité, mesurée par un test, de l'application d'une habileté. Dans l'exemple fictif illustré dans l'encadré de gauche de la figure 5, le coefficient de corrélation de 0,86 indique une forte association, une corrélation parfaite étant de 1,00. Les études d'association diffèrent des études expérimentales de plusieurs manières. Dans les études d'association, on s'intéresse à différents participants mais il n'y a pas de groupes. La variable indépendante est parfois appelée « prédicteur » et la variable dépendante « critère ». Dans les études expérimentales, on s'attache à recruter des participants aussi similaires que possible, qui sont ensuite répartis dans deux groupes dont l'un est soumis à une intervention (groupe I) et dont l'autre sert de groupe contrôle (groupe C) non soumis à l'intervention. On mesure ensuite le résultat. La variable indépendante représente la cause et la variable dépendante l'effet ou le résultat. On considère que l'intervention a eu un effet réel lorsqu'on met en évidence une différence de

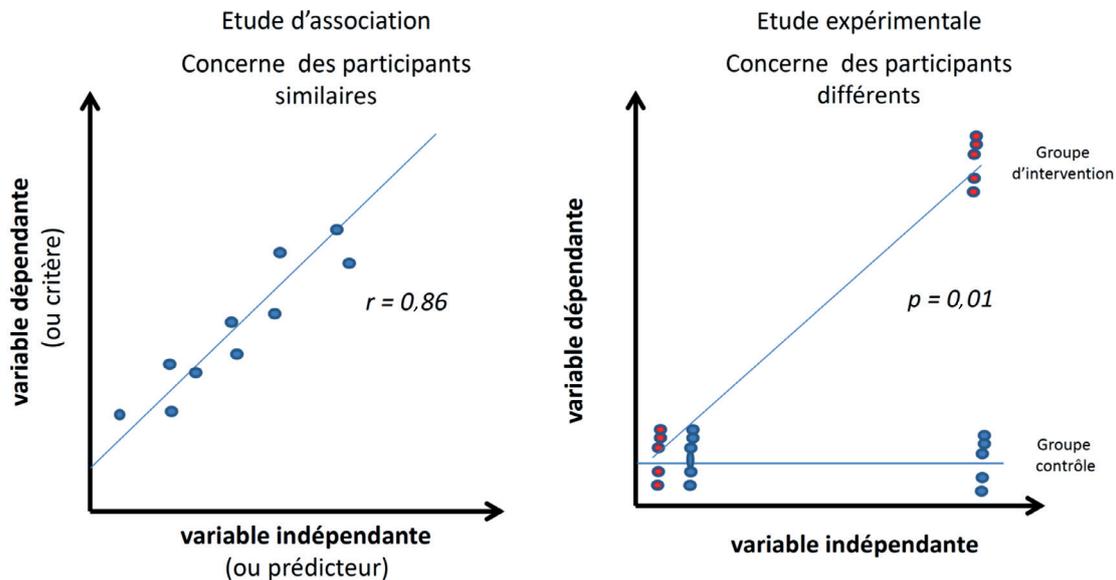


Fig. 5. différence entre les études d'association et les études expérimentales. La variable indépendante est la cause/l'intervention et la variable dépendante est l'effet/le résultat.

résultat entre les groupes, avec une valeur « p » indiquant la probabilité que cette différence ne soit pas due au hasard.

Dans les études observationnelles, il y a souvent plus d'une variable indépendante, ce qui nécessite des méthodes d'analyse plus sophistiquées – c'est-à-dire des analyses de régression multivariée – pour déterminer la contribution de chacune d'elles au résultat. L'analyse de variance est donc fréquemment utilisée dans les études de cohorte et les études cas-témoins^[36, 32, 34]. Ce guide introductif n'ira pas plus loin à ce sujet, qui fera l'objet d'autres guides AMEE de la série « Recherche ».

Le défi

Le défi des études observationnelles est que le chercheur ne peut pas contrôler les biais liés à des variables confondantes. Un autre défi concerne le choix de méthodes d'analyse et d'interprétation des données^[37]. Il est dès lors fortement recommandé de consulter un statisticien dès le début de la conception de l'étude. Enfin, les relations mises en évidence dans les études observationnelles ne peuvent pas être considérées comme démontrant un lien de causalité ;

tout au plus peuvent-elles le suggérer, en raison, notamment, des nombreux facteurs, connus ou non, susceptibles d'influencer le résultat. Les chercheurs devront s'appuyer sur un solide cadre théorique pour fournir des explications soigneusement argumentées au sujet des fondements de leurs attentes concernant une éventuelle relation. Par ailleurs, ce type d'études nécessite des connaissances approfondies en matière de gestion et d'interprétation des données. On notera cependant que si les relations observées ne peuvent être considérées comme des preuves directes d'une relation causale, elles peuvent être utilisées pour guider des travaux ultérieurs recourant aux RCT ou à d'autres approches expérimentales contrôlées. Enfin, l'établissement d'associations entre processus pédagogiques, résultats pédagogiques et qualité des soins nécessite des bases de données contenant à la fois des données de nature clinique et des données de nature pédagogique. Face à l'absence de telles banques de données, certains ont appelé de leurs vœux la création de banques de données pédagogiques recueillant de manière prospective des données concernant les apprenants dans le temps et à travers plusieurs institutions, pour faciliter la conduite d'études observationnelles^[15].

Les études translationnelles

Le concept de recherche translationnelle provient du monde biomédical et répond à la crainte que les découvertes scientifiques ne soient pas transposées à la pratique clinique, où elles peuvent bénéficier aux patients et à la population générale^[38]. Du laboratoire de sciences fondamentales à la pratique clinique, la route est longue et comprend essentiellement trois grandes étapes. La première est de traduire les découvertes effectuées en sciences fondamentales en études cliniques orientées vers les patients. La deuxième est de synthétiser les connaissances issues d'études cliniques empiriques et d'élaborer des recommandations de bonne pratique, étape encore appelée « traduction des connaissances »^[39]. Lors de la troisième étape, il s'agit d'implanter les nouvelles connaissances en pratique clinique et d'évaluer si, comment et pourquoi cela fonctionne dans divers contextes complexes de la réalité de terrain, c'est-à-dire, de mener des études d'efficacité. Contrairement à une conception du processus translationnel comme une trajectoire unidirectionnelle du laboratoire au lit du patient, les relations quant au transfert de connaissances issues de la recherche entre les sciences fondamentales et la clinique sont de plus en plus souvent perçues comme cycliques ou multidirectionnelles^[4, 40].

La recherche en éducation médicale et la recherche translationnelle peuvent être mutuellement contributives de diverses manières. Tout d'abord, la recherche en éducation médicale peut s'intéresser à l'enseignement et l'apprentissage de concepts liés à la recherche translationnelle dans le domaine biomédical. En effet, ceux-ci peuvent être intégrés aux cursus médicaux pré et postgradués. Par ailleurs, la recherche en éducation médicale peut éclairer la compréhension des processus d'élaboration et d'implantation des recommandations de bonne pratique, en s'intéressant au développement professionnel continu des médecins. Elle peut en effet permettre de mieux comprendre comment les professionnels de santé cherchent et assimilent de nouveaux savoirs pour, *in fine*, modifier leurs comportements et leurs pratiques^[41, 42]. Troisièmement, la recherche translationnelle dans le domaine de l'éducation médicale

peut être comprise comme une recherche visant à comprendre comment l'apprentissage en amphithéâtre ou en milieu simulé se traduit en compétences dans la pratique réelle et, *in fine*, influence l'état de santé des patients^[43]. Enfin, la recherche translationnelle en éducation médicale peut également être conçue comme visant à comprendre comment les principes issus des sciences fondamentales et appliquées en éducation et en didactique peuvent être utilisés dans la pratique pédagogique quotidienne en sciences de la santé.

Les études translationnelles en éducation médicale seront abordées plus en détail dans les trois parties suivantes : création de savoirs synthétiques, implantation des connaissances et études d'efficacité.

Création de savoirs synthétiques (revues de littérature)

Les études visant à créer des savoirs synthétiques portent sur les recherches antérieures et constituent une littérature scientifique dite secondaire. Il s'agit des revues de la littérature, telles que les revues systématiques, les revues réalistes, les revues BEME (« *Best Evidence Medical Education* » ou « éducation médicale fondée sur les meilleures données probantes »). Comme toute recherche, une revue systématique doit s'inscrire dans un cadre théorique ou conceptuel, partir d'une ou plusieurs questions de recherche, utiliser une approche systématique de recueil des données, des critères d'inclusion et d'exclusion explicites et une approche systématique d'analyse et d'interprétation des données.

Les revues systématiques ont généralement une orientation quantitative et cherchent à compiler les données concernant les effets d'une intervention pour en dériver une taille d'effet moyenne. Un certain nombre de revues systématiques sont publiées dans les revues d'éducation médicale et les revues biomédicales. Les revues quantitatives se heurtent cependant aux difficultés liées à la variance observée entre études quant aux éléments SPICOT, abordés dans la partie sur les études expérimentales. Si l'on ne tient pas compte des différences entre études sur le plan des éléments SPICOT, on risque de colliger des études dont les résultats vont dans des directions opposées.

Par exemple, le choix des temps d'évaluation peut avoir une influence majeure sur les résultats. Il n'est toutefois pas aisé d'en tenir compte car les éléments SPICOT sont souvent mal décrits dans les études expérimentales. Dès lors, il ne faut pas s'étonner que bon nombre de revues systématiques quantitatives ne parviennent à objectiver que des effets modestes, voire des résultats peu concluants. Les revues systématiques sont néanmoins précieuses pour le chercheur qui envisage d'étudier une thématique donnée car elles lui procureront une excellente vue d'ensemble des recherches antérieures, ainsi qu'une mine de références utiles.

Les problèmes liés aux revues systématiques quantitatives tiennent également au manque d'attention portée aux cadres théoriques ou conceptuels. Ceux-ci sont en revanche une préoccupation majeure des revues dites réalistes, qui replacent les données quantitatives empiriques dans un cadre théorique ou conceptuel. Ce type de revue explore précisément des théories qui permettent de rendre compte des différences observées d'une étude à l'autre^[43-45].

Les revues BEME sont tantôt des revues systématiques quantitatives tantôt des revues de type réaliste. Elles ont souvent une visée avant tout pratique, fournissant des recommandations quant aux meilleures données disponibles et leur contribution potentielle à la pratique éducative.

Les revues narratives et critiques sont moins exhaustives. Elles cherchent avant tout à obtenir la quintessence des recherches antérieures et à identifier les problèmes qui méritent d'être explorés. Ce type de revue est généralement mené au début d'un projet de recherche et sert à définir le contexte et le cadre théorique ou conceptuel de l'étude^[46].

Implantation des connaissances

Au début de cette partie consacrée à la recherche translationnelle, nous avons souligné que l'implantation de recommandations de bonne pratique et l'adoption de la pratique fondée sur les faits (*evidence-based practice*) sont des thèmes extrêmement porteurs dans le domaine biomédical. Les recommandations de bonne pratique fondées sur des revues systématiques sont

des outils-clés dans l'implantation des connaissances dans le domaine biomédical. La formulation de recommandations du même type dans le domaine de l'éducation médicale ne va pas de soi, d'une part parce qu'il y a peu de recommandations basées sur les données probantes et, d'autre part, parce que bon nombre d'enseignants ne connaissent pas la littérature en éducation médicale. Les guides de l'AMEE adoptent une approche pragmatique et visent à stimuler l'esprit critique et la réflexion, tout en fournissant des conseils et un soutien pratiques. Ils sont en quelque sorte inspirés des données probantes, plutôt que fondés sur celles-ci. Les revues BEME partagent davantage de points communs avec les recommandations de bonne pratique du domaine médical. Toutefois, la complexité de l'éducation médicale et l'interaction de nombreux facteurs (figure 1) rendent la création de recommandations universelles quasiment impossible en raison de l'immense diversité des caractéristiques, opportunités et contraintes locales^[2].

De nombreux enseignants ressentent une certaine frustration quant aux modes d'enseignement traditionnels pratiqués dans leurs institutions et s'évertuent à implanter et à perpétuer des pratiques pédagogiques innovantes. Organiser des cours, des ateliers et des séminaires permet de disséminer les connaissances issues de l'éducation médicale fondée sur les données probantes. Il s'agit là d'une stratégie répandue mais, malgré l'existence de preuves d'un effet bénéfique de cours et d'initiatives de développement professoral, on en sait encore trop peu sur les questions du « pourquoi » et du « comment » de l'efficacité de ces stratégies^[8, 47, 48]. Plutôt que d'adopter une approche directive et didactique, une autre manière de promouvoir le changement est de se joindre aux parties prenantes (*stakeholders*) pour élaborer des recommandations^[49]. Le fait d'impliquer les parties prenantes dans le processus est l'un des indicateurs de qualité récemment validés concernant le processus d'élaboration et d'implantation des recommandations de bonne pratique clinique^[50]. Bon nombre de ces indicateurs pourraient également s'appliquer à l'élaboration de recommandations de bonne pratique en éducation médicale. Par ailleurs, il existe un éventail de cadres théoriques ou conceptuels concernant la dissémination des connaissances, qui

pourraient être appliqués à la traduction des connaissances^[51,4]. Enfin, transformer des innovations pédagogiques en projets de recherche pourrait s'avérer une stratégie payante dans le domaine médical^[1], en raison, d'une part, de l'ouverture induite par la question de recherche et, d'autre part, de l'implication des parties prenantes. Nous aborderons cela plus en détail dans la partie suivante.

Les études d'efficience

Alors que les études expérimentales cherchent à savoir si une intervention fonctionne dans des conditions contrôlées, les études d'efficience cherchent à savoir ce qui fonctionne, comment et pour qui, dans des conditions réelles. Dans le domaine biomédical, il y a une prise de conscience croissante de la multiplicité des facteurs déterminant l'application des connaissances issues des sciences fondamentales dans la pratique clinique, ainsi que de la nécessité d'approches plurielles pour évaluer les interventions^[52, 53]. Ces phénomènes constituent l'objet des études d'efficience, qui examinent les interventions complexes, plutôt que les traitements uniques simples. La complexité relève du nombre d'éléments et de la flexibilité de l'intervention, de la diversité des comportements attendus des utilisateurs, du nombre de groupes ou de niveaux organisationnels impliqués et de la variance des résultats^[52]. Les études d'efficience ont généralement une orientation très pragmatique et cherchent des réponses pertinentes pour les parties prenantes utiles à la prise de décision^[54]. Ces études comprennent diverses mesures de processus et de résultats, afin d'évaluer les effets sur les individus, les systèmes et les organisations, y compris sur le plan du rapport coût-efficacité. L'idée d'évaluer des effets à plusieurs niveaux (réaction, apprentissage, comportement et organisation) est bien connue dans le domaine de l'éducation médicale, grâce au modèle de Kirkpatrick^[55]. Cependant, la nouvelle tendance en matière d'études d'efficience est d'élargir leur orientation de la simple évaluation à l'étude de questions d'élaboration, de faisabilité et d'implantation. L'élaboration comprend l'identification des données et de la théorie sous-jacente pour modéliser les processus et les résultats. La faisabilité comprend la conduite des processus, la prise en compte des questions de

recrutement et de rétention et la détermination des tailles d'effet de l'étude principale à large échelle. Les questions d'implantation touchent à la dissémination, la surveillance et le monitoring, ainsi qu'au suivi à long terme. La conduite des quatre composantes des études d'efficience, à savoir l'élaboration, la faisabilité, l'implantation et l'évaluation, doit être comprise comme un processus itératif d'allers-retours plutôt qu'un processus par étapes^[52]. De plus amples détails concernant ce cadre sont présentés sur le site du *Medical Research Council* (MRC) britannique (www/mrc.ac.uk/complexinterventionsguidance).

Bien que les aspects homologues de l'élaboration et de l'évaluation d'interventions complexes en recherche en éducation médicale puissent ressembler à ces initiatives dans le domaine biomédical, de nombreux domaines en éducation médicale ne se prêtent pas à la diversité des approches de recherche recommandées par le MRC. Les principes et la rigueur préconisés sont néanmoins tout aussi pertinents en éducation médicale et des aspects tels que le rapport coût-efficacité devraient faire l'objet d'une plus grande attention en recherche en éducation médicale. Plusieurs concepts liés aux études d'efficience sont proches de ceux associés à la recherche dite de conception itérative (*design-based research*), avec ses éléments-clés que sont « la conception, la réalisation, l'implantation et l'évaluation ». Les études d'efficience, comme les études de « conception itérative », peuvent avoir une large envergure. Les deux approches sont similaires quant aux recommandations d'impliquer de multiples sites et donc, de nombreux participants dans leurs travaux. Néanmoins, une partie des résultats émergents peut éclairer la compréhension de ce qui fonctionne, des causes et des obstacles, des cibles privilégiées et peuvent ainsi être publiés sous forme d'articles. Comme cela a déjà été précisé dans l'introduction de ce guide, les études d'efficience se distinguent de la simple évaluation ou de l'audit par leur référence à un cadre théorique ou conceptuel, par l'explicitation de questions de recherche et la mise en œuvre d'une approche rigoureuse. Actuellement, la littérature comprend davantage de rapports concernant des protocoles d'études d'efficience que de résultats à proprement parler.

Conclusions

Ce guide de l'AMEE propose une introduction à la recherche en éducation médicale. Il aborde quelques principes de base concernant la recherche dans le domaine de l'éducation médicale et présente une vue d'ensemble des différentes approches de recherche possibles. Nous espérons qu'ayant lu cette introduction, le lecteur ne verra plus la recherche appliquée et la recherche théorique comme deux pôles opposés mais, au contraire, qu'il saisira la complémentarité de ces deux approches imbriquées et interdépendantes, partageant le même objectif d'accroître les savoirs en éducation médicale. Les guides suivants de l'AMEE aborderont les diverses approches de recherche plus en détail.

Conflits d'intérêts

Les auteurs ne signalent aucun conflit d'intérêts concernant cet article.

Contribution

Charlotte Ringsted est professeur d'éducation médicale et directrice du centre d'éducation clinique de l'Université de Copenhague et Région Capitale, faculté des sciences de la santé, Danemark.

Brian Hodges est un enseignant-chercheur, directeur du Centre Wilson pour la recherche en éducation médicale, Université de Toronto, faculté de médecine, et vice-président « éducation » du University Health Network, Canada.

Albert Scherpier est professeur d'éducation médicale, directeur scientifique de l'institut pour l'éducation, faculté des sciences de la santé, médecine et sciences de la vie, et vice-doyen à l'éducation de l'Université de Maastricht, Pays-Bas.

Références

- Grant J, Gale R. Changing medical education. *Med Educ* 1989;23:252-7.
- Regehr G. It's NOT rocket science: Rethinking our metaphors for research in health professions education. *Med Educ* 2010;44:31-39.
- Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, Tyrer P. Framework for design and evaluation of complex interventions to improve health. *BMJ* 2000;321:694-6.
- Ward V, House A, Hamer S. Developing a framework for transferring knowledge into action: A thematic analysis of the literature. *J Health Serv Res Policy* 2009;14:156-64.
- Bunniss S, Kelly DR. Research paradigms in medical education research. *Med Educ* 2010;44:358-66.
- Carpiano RM, Daley DM. A guide and glossary on post-positivist theory building for population health. *J Epidemiol Commun Health* 2006; 60:564-70.
- Bordage G. Conceptual frameworks to illuminate and magnify. *Med Educ* 2009;43:312-319.
- Mann KV. The role of educational theory in continuing medical education: Has it helped us? *J Contin Educ Health Prof* 2004;24(Suppl 1):S22-S30.
- Bleakley A. Broadening conceptions of learning in medical education: The message from teamworking. *Med Educ* 2006;40:150-7.
- Patel VL, Yoskowitz NA, Arocha JF. Towards effective evaluation and reform in medical education: A cognitive and learning sciences perspective. *Adv Health Sci Educ Theory Pract* 2009;14:791-812.
- Patel VL, Yoskowitz NA, Arocha JF, Shortliffe EH. Cognitive and learning sciences in biomedical and health instructional design: A review with lessons for biomedical informatics education. *J Biomed Inform* 2009;2:176-97.
- Mann KV. Theoretical perspectives in medical education: Past experience and future possibilities. *Med Educ* 2011;45:60-8.
- Haig A, Dozier M. BEME Guide no 3: Systematic searching for evidence in medical education—Part 1: Sources of information. *Med Teach* 2003;25:352-63.
- Haig A, Dozier M. BEME Guide no. 3: Systematic searching for evidence in medical education—Part 2: Constructing searches. *Med Teach* 2003;25:463-84.
- Cook DA, Andriole DA, Durning SJ, Roberts NK, Triola MM. Longitudinal research databases in medical education: Facilitating the study of educational outcomes over time and across institutions. *Acad Med* 2010;85:1340-6.
- Cook DA, Beckman TJ, Bordage G. A systematic review of titles and abstracts of experimental studies in medical education: Many informative elements missing. *Med Educ* 2007;41:1074-81.

17. Todres M, Stephenson A, Jones R. Medical education research remains the poor relation. *BMJ* 2007;335:333-5.
18. Pope C, Mays N. Reaching the parts other methods cannot reach: An introduction to qualitative methods in health and health services research. *BMJ* 1995;311:42-5.
19. Greenhalgh T, Taylor R. Papers that go beyond numbers (qualitative research). *BMJ* 1997;315:740-3.
20. Kuper A, Reeves S, Levinson W. An introduction to reading and appraising qualitative research. *BMJ* 2008;337:a288.
21. Reeves S, Kuper A, Hodges BD. Qualitative research methodologies: Ethnography. *BMJ* 2008;337:a1020.
22. Hodges BD, Kuper A, Reeves S. Discourse analysis. *BMJ* 2008;337:a879.
23. Lingard L, Albert M, Levinson W. Grounded theory, mixed methods, and action research. *BMJ* 2008;337:a567.
24. Kvale S. *InterViews: An introduction to qualitative research interviewing*. Thousand Oaks (CA): Sage Publications, 1996.
25. Kuper A, Lingard L, Levinson W. Critically appraising qualitative research. *BMJ* 2008;337:a1035.
26. Swanson DB, Case SM, Nungester RJ. Validity of NBME Part I and Part II in prediction of Part III performance. *Acad Med* 1991;66:S7-9.
27. Ramsey PG, Carline JD, Inui TS, Larson EB, LoGerfo JP, Wnrich MD. Predictive validity of certification by the American Board of Internal Medicine. *Ann Int Med* 1989;110: 719-26.
28. Streiner DL, Norman GR. *Health Measurement Scales: A practical guide to their development and use*. 4th ed. Oxford: Oxford University Press, 2008.
29. Haynes R. Forming research questions. *J Clin Epidemiol* 2006;59:881-6.
30. Wulf G, Shea CH. Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychon Bull Rev* 2002;9:185-211.
31. Illeris K. *The three dimensions of learning*. Copenhagen: Roskilde University Press/Leicester: NIACE, 2004.
32. Norman GR, Wenghofer E, Klass D. Predicting doctor performance outcomes of curriculum interventions: Problem-based learning and continuing competence. *Med Educ* 2008;42:794-9.
33. Tamblyn R, Abrahamowicz M, Dauphinee WD, Hanley JA, Norcini J, Girard N, Grand'Maison P, Brailovsky C. Association between licensure examination scores and practice in primary care. *JAMA* 2002;288:3019-26.
34. Cadieux G, Abrahamowicz M, Dauphinee D, Tamblyn R. Are physicians with better clinical skills on licensing examinations less likely to prescribe antibiotics for viral respiratory infections in ambulatory care settings? *Med Care* 2011;49:156-65.
35. Papadakis MA, Hodgson CS, Teherani A, Kohatsu ND. Unprofessional behavior in medical school is associated with subsequent disciplinary action by a state medical board. *Acad Med* 2004;79:244-9.
36. Teherani A, Hodgson CS, Banach M, Papadakis MA. Domains of unprofessional behavior during medical school associated with future disciplinary action by a state medical board. *Acad Med* 2005;80:S17-S20.
37. Colliver JA, Markwell SJ, Verhulst SJ, Robbs RS. The prognostic value of documented unprofessional behavior in medical school records for predicting and preventing subsequent medical board disciplinary action: The Papadakis studies revisited. *Teach Learn Med* 2007;19:213-5.
38. Zerhouni EA. Translational and clinical science – Time for a new vision. *N Engl J Med* 2005;353: 1621-3.
39. Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, Robinson N. Lost in knowledge translation: Time for a map? *J Contin Educ Health Prof* 2006;26:13-24.
40. Rubio DM, Schoenbaum EE, Lee LS, Scheingart DE, Marantz PR, Anderson KE, Platt LD, Baez A, Esposito K. Defining translational research: Implications for training. *Acad Med* 2010;85:470-5.
41. Slotnick HB. How doctors learn: Physicians' self-directed learning episodes. *Acad Med* 1999;74: 1106-17.
42. Davis N, Davis D, Bloch R. Continuing medical education: AMEE Education Guide No. 35. *Med Teach* 2008;30:652-66.
43. McGaghie WC. Medical education research as translational science. *Sci Transl Med* 2010;2:19cm8.
44. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review – A new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy* 2005; 10(Suppl 1):21-34.
45. Wong G, Greenhalgh T, Pawson R. Internet-based medical education: A realist review of what works, for whom and in what circumstances. *BMC Med Educ* 2010;10:12.

46. Eva KW. On the limits of systematicity. *Med Educ* 2008;42:852-3.
47. Steinert Y, Mann K, Centeno A, Dolmans D, Spencer J, Gelula M, Prideaux D. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Med Teach* 2006;28:497-526.
48. Steinert Y, McLeod PJ, Boillat M, Meterissian S, Elizov M, Macdonald ME. Faculty development: A “field of dreams”? *Med Educ* 2009;43:42-9.
49. Grol R, Grimshaw J. From best evidence to best practice: Effective implementation of change in patients’ care. *Lancet* 2003;362:1225-30.
50. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, Fervers B, Graham ID, Hanna SE, Makarski J. Development of the AGREE II, part 1: Performance, usefulness and areas for improvement. *CMAJ* 2010;182:1045-52.
51. Estabrooks CA, Thompson DS, Lovely JJ, Hofmeyer A. A guide to knowledge translation theory. *J Contin Educ Health Prof* 2006;26:25-36.
52. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: The new Medical Research Council guidance. *BMJ* 2008;337:a1655.
53. Ward VL, House AO, Hamer S. Knowledge brokering: Exploring the process of transferring knowledge into action. *BMC Health Serv Res* 2009;9:12.
54. Tunis SR, Benner J, McClellan M. Comparative effectiveness research: Policy context, methods development and research infrastructure. *Stat Med* 2010;29:1963-76.
55. Kirkpatrick D. *Evaluating training programs*. San Francisco (CA): Berrett-Kochler, 1998.

Correspondance et offprints : C. Ringsted, Centre for Clinical Education, University of Copenhagen and Capital Region, Rigshospitalet, 5404, Blegdamsvej 9, 2100 Copenhagen, Denmark.
Mailto: charlotte.ringsted@rh.regionh.dk