

La recherche scientifique en éducation: Quels critères de qualité?

JEAN-MARIE DE KETELE

Abstract: *Which criteria make the quality of the scientific research process, whose purpose is education or training? How ensure scientific relevance, validity, trustworthiness in research?*

The purpose of this article is attempting to formalize a set of coherent and essential criteria to any scientific educational research, considering three aspects: the conceptualization, the collection and the processing of data, the formulation of conclusions.

By examining some current issues away from the epistemological and methodological sectarianism, the author confirms the first function of any scientific activity: producing intelligibility.

Resumé: *À quels critères soumettre la qualité du processus des recherches scientifiques dont l'objet est l'éducation ou la formation? Comment, dans ces recherches, assurer la pertinence scientifique, la validité, la fiabilité?*

L'objet de cet article est bien de formaliser un ensemble de critères cohérents et incontournables pour toute recherche scientifique en éducation et qui sont à examiner selon trois aspects essentiels: la conceptualisation, le recueil et le traitement des données, l'énonciation des conclusions.

En examinant quelques questions actuelles au de là de sectarismes épistémologiques et méthodologiques, l'auteur ne renonce à confirmer la fonction première de toute entreprise scientifique: produire de l'intelligibilité.

Mots clés: *éducation, recherche, critères scientifiques, qualité, intelligibilité.*

1. La fonction commune de toute recherche

La recherche a pour fonction de produire de nouvelles connaissances. Il existe de nombreuses façons de produire de la connaissance. Le commissaire de police qui enquête sur un crime cherche à identifier le meurtrier et les raisons de son crime (connaissances nouvelles). Le romancier produit également de la connaissance lorsqu'il dépeint la psychologie de ces personnages. Le poète approche souvent une vérité que la science ne peut

atteindre. A leur façon, le commissaire, le romancier et le poète sont des chercheurs, chacun dans leur domaine et avec leur stratégie, et n'ont pas moins d'importance que le scientifique. Quand peut-on dire alors que le chercheur conduit une recherche scientifique? La recherche est dite scientifique s'il s'agit d'un processus ayant pour fonction de produire des connaissances scientifiques nouvelles provisoires à l'aide d'activités méthodiques, rigoureuses et vérifiables, et soumises à la légitimation de la communauté scientifique. Plusieurs caractéristiques de cette définition méritent d'être rappelées: (1) les connaissances scientifiques sont toujours provisoires, car susceptibles d'être contredites même si, au moment où elles sont énoncées, elles forment un ensemble cohérent avec les connaissances scientifiques disponibles; (2) la recherche ne peut être considérée comme scientifique si elle n'est pas sujette au débat scientifique par les pairs actuels et futurs.

L'objet de cet article est bien de tenter de formaliser un ensemble de critères cohérents permettant de soumettre la qualité du processus des recherches scientifiques dont l'objet est l'éducation ou la formation. Dans ce travail, nous nous limitons aux recherches scientifiques qui obéissent aux deux conditions suivantes: (1) elles portent sur des pratiques éducatives ou de formation ou sur les conditions qui les influencent ou sur leurs effets (elles peuvent se limiter à l'un ou l'autre de ses aspects ou les englober tous); (2) elles donnent lieu à la production de données empiriques, qu'elles soient d'ordre quantitatif ou qualitatif (ou les deux). Nous excluons donc ici de notre réflexion les recherches exclusivement spéculatives, c'est-à-dire sans production de données empiriques; nous n'oublions cependant pas que la réflexion spéculative est indispensable et constitue une composante essentielle du processus de recherche scientifique car elle permet de fonder le cadre problématique de la recherche et la production des données qui en résultera.

Il nous paraît aussi important de souligner d'emblée que «l'opposition entre le quantitatif et le qualitatif n'est pas le bon débat, malgré le fait que les farouches partisans de chacune des approches ont des arguments qui méritent un examen attentif» (De Ketele, Maroy, 2006, 219). Le vrai débat est celui de la qualité de la recherche: permet-elle de produire des connaissances nouvelles avec rigueur, c'est-à-dire en utilisant des démarches, quantitatives ou qualitatives, dont la valeur peut être vérifiée par la communauté scientifique (actuelle et future).

Notons aussi que si la fonction principale de toute recherche est de produire de la connaissance, celle-ci peut servir des visées multiples: (1) la connaissance est formulée sous forme de loi qui prendra une place cohérente dans un ensemble de lois ou dans un “corps de connaissance” (cette visée caractérise le plus souvent les recherches des disciplines qui sont au fondement des sciences de l’éducation); (2) la connaissance est explicitée sous forme de modélisation qui permet de mettre en évidence le jeu des relations complexes entre plusieurs variables, dimensions ou facteurs intervenant dans le phénomène étudié (d’où l’intérêt actuel des chercheurs en éducation pour les modèles des pistes causales, des modèles multiniveaux, etc.); (3) la connaissance permet de décrire de façon rigoureuse un phénomène que l’on avait du mal à appréhender jusqu’à présent (ex.: la maltraitance; voir Cadei, 2006); (4) la connaissance permet de mettre en évidence une démarche d’analyse pertinente qui donne du sens à un ensemble d’informations récoltées avec méthode dans un processus “idiographique” (étude de cas, récit de vie...); (5) la connaissance est un produit évaluatif mis au service de la prise de décision et permettant des raisonnements du type “si... alors” (si tel objectif et tels critères sont posés au départ... alors on peut dire que le contexte décrit de telle façon a permis d’obtenir tels effets; par contre si tel autre objectif et tels critères sont posés au départ, ce même contexte décrit de telle façon a permis d’obtenir tels autres effets). Ces quelques exemples montrent bien que la recherche conduite avec une démarche scientifique peut remplir des visées multiples, même si la fonction commune est de produire de la connaissance. Nous défendons donc la thèse qu’il existe avant tout des critères communs à toute recherche pour en évaluer la qualité.

2. Trois critères fondamentaux et aspects à prendre en considération

Quelle que soit la démarche adoptée, quantitative ou/et qualitative, trois critères de qualité sont incontournables: la pertinence, la validité et la fiabilité. Chacun de ces trois critères sont à examiner selon trois aspects: la conceptualisation, le recueil et le traitement des données, l’énonciation des conclusions.

2.1. La pertinence

Trop peu souvent mentionnée car moins technique, la pertinence est la qualité première de toute recherche. Elle correspond à la question centrale “*Est-ce que je ne me trompe pas de...?*”. La recherche est d’autant plus pertinente que je ne me trompe pas de conceptualisation (premier aspect à examiner), de recueil et de traitement des données (deuxième aspect), d’énonciation des conclusions (aspect).

2.1.1. La pertinence scientifique

Ne pas se tromper de conceptualisation suppose que l’on examine avec attention les points suivants: bien poser le problème de la recherche, faire reposer la recherche sur un cadre conceptuel adéquat et sur un cadre problématique cohérent.

Bien poser le problème de la recherche, c’est tout d’abord ne pas se tromper d’objet de recherche. Ainsi une revue critique de la littérature scientifique permet d’éviter de refaire moins bien ce qui a déjà été bien fait, à moins que l’analyse critique permette de montrer que d’autres angles d’attaque sont possibles et permettraient de produire une connaissance nouvelle ou, à tout le moins, d’infirmar ou de contester le bien fondé des résultats de recherches antérieures. Bien poser le problème, c’est aussi ne pas s’embarquer dans un processus de recherche alors qu’un peu de bon sens et de réalisme montreraient que le chercheur n’est pas en mesure de produire des données quantitatives ou qualitatives pertinentes pour apporter une réponse adéquate à la question posée. Bien poser le problème, c’est aussi se souvenir que tout chercheur a dès le départ des hypothèses. Un chercheur qui s’en défend n’est pas crédible et devient même dangereux. Ne pas s’en rendre compte, et donc ne pas les expliciter, est aussi périlleux que de vouloir à tout prix démontrer ses hypothèses. Adopter une attitude scientifique consiste à expliciter ses hypothèses, à les considérer comme provisoires et susceptibles d’être infirmées, à tout faire sur le plan méthodologique pour les infirmer et être prêt à poser de nouvelles hypothèses. La question de savoir s’il faut nécessairement poser ses hypothèses avant de recueillir ses données dépend du type de recherche: on voit mal comment créer un dispositif expérimental (au sens des “*experimental designs*” de Campbell et Stanley, 1966) sans énoncer de façon précise l’hypothèse à vérifier; par contre, d’autres types de recherches produiront des données quantitatives et qualitatives dans le

but explicite de faire naître des hypothèses pertinentes qui permettront de poursuivre la récolte des données ou les modalités de traitement de celles-ci de telle façon que les hypothèses générées puissent être mises à l'épreuve.

Faire reposer la recherche sur un cadre conceptuel adéquat consiste à identifier les concepts nécessaires et suffisants (ni trop, ni trop peu), de les définir avec soin, de les distinguer des concepts proches et d'organiser les concepts retenus. Identifier les concepts nécessaires est d'autant plus difficile que ce sont des concepts abstraits utilisés dans la vie courante. Ici encore, une bonne revue critique de la littérature scientifique s'avère importante. Ainsi, à titre d'exemple, nous savons grâce à des travaux comme ceux de Bandura (2003) ou la revue critique de Viau (2004) qu'il importe de distinguer les concepts de motivation et d'engagement: s'il est fréquent de s'engager sur la base d'une motivation forte, il n'est pas rare de s'engager dans l'étude de tel cours sans être vraiment motivé par ce cours, comme il n'est pas rare de ne pas s'y engager malgré une bonne motivation. Des métaphores sont parfois utiles pour bien faire apparaître ce qui permet de bien distinguer deux concepts. Ainsi, la motivation est de l'ordre de "l'énergie" (le carburant qui permet à une voiture de se déplacer, du moins potentiellement car cela ne suffit pas; il faut un moteur en état de marche et un conducteur sachant conduire), l'engagement est de l'ordre du "travail" au sens physique du terme c'est-à-dire d'une énergie transformée en acte. Une voiture peut réellement se déplacer sans l'aide du carburant lorsqu'elle est mue par une autre force, comme celle de la gravité dans une pente ou toute autre force suffisante.

Faire reposer la recherche sur un cadre problématique pertinent consiste à identifier les variables (dimensions ou facteurs selon les cas) nécessaires et suffisantes pour tenter de répondre à la question posée. Le cadre conceptuel de la recherche est souvent à l'origine de l'élaboration du cadre problématique. Ainsi, Viau (*Ibidem*) organise les variables qu'il retient pour étudier la motivation en contexte scolaire comme suit (voir figure 1).

Très souvent, l'élaboration du cadre problématique suppose l'identification et l'organisation de variables qui ne sont pas toutes induites par le cadre conceptuel de départ. Ici encore, la revue critique de la littérature scientifique s'avère précieuse. Dans certains domaines, il existe même des grilles de base (dégagées d'un grand nombre de recherches) pour aider le chercheur à élaborer son cadre problématique. Ainsi, par exemple, beaucoup de recher-

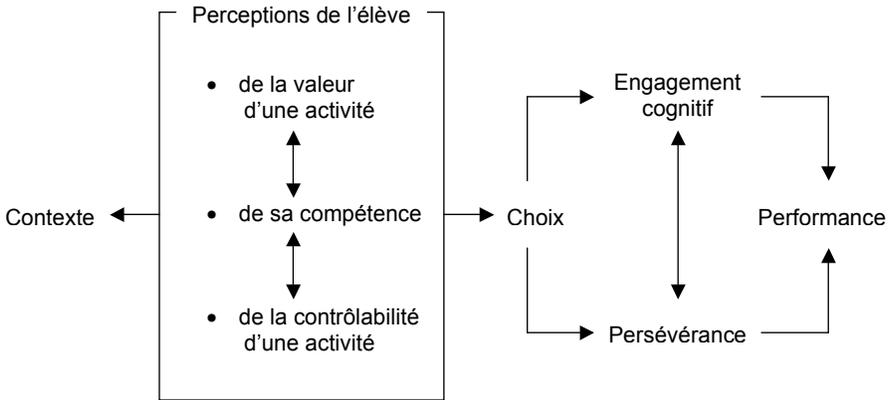


Figure 1: Modèle de motivation en contexte scolaire (Viau, 2004, 32).

ches prenant comme objet l'efficacité en contexte scolaire ont construit leur cadre problématique en s'appuyant sur le modèle de Creemers (1994 in Paquay, Crahay & De Ketele, 2006), et présenté dans la figure 2.

Les modèles théoriques sont donc d'une très grande utilité. Il doivent cependant être adaptés au problème contextualisé par le chercheur, éventuellement contestés ou complétés sur certains points, souvent également réduits dans le sens où les conditions de la recherche imposent que l'on se penche sur une partie du cadre problématique général pour étudier plus en profondeur certaines variables et leurs relations. Notons que la représentation de son cadre problématique sous forme de schéma est souvent la meilleure façon de clarifier ses hypothèses.

2.1.2. La pertinence méthodologique

La pertinence méthodologique consiste à ne pas se tromper de recueil et de traitement des données. Le cadre problématique établi au départ ou en cours de route ou ajusté après un premier type de recueil de données est évidemment une base indispensable pour identifier les informations à recueillir, les modalités de recueil et de traitement des données.

Les informations recueillies ne sont pas pertinentes si elles ne sont pas reliées au cadre problématique posé. Ainsi, on voit encore de trop nom-

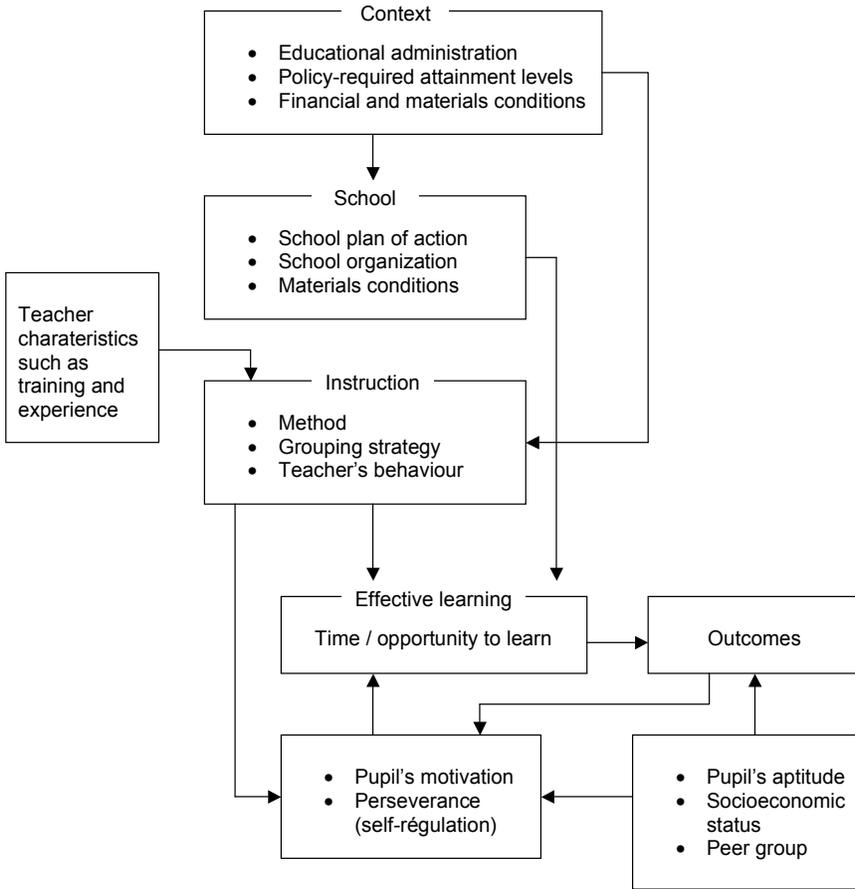


Figure 2: Modèle d'apprentissage en contexte scolaire de Creemers (1994) in: Paquay, Crahay, De Ketele (2009)

breux questionnaires dont les questions ne sont pas explicitement mises en relation avec les variables ou dimensions ou facteurs identifiés dans le cadre problématique ou qui s'écartent des définitions présentées dans le cadre conceptuel. S'il va de soi que la présentation du questionnaire à la personne enquêtée ne met pas en évidence ces mises en relation (ce qui dans certains cas introduirait un biais dans les réponses), le rapport de recherche doit les faire apparaître.

Bien choisir les modalités pertinentes du recueil des informations est une autre question importante, car chaque outil de recueil a en soi certains avantages et certaines limites qui peuvent d'ailleurs être amplifiées par les modalités d'utilisation. Ainsi, le questionnaire d'enquête peut être un outil adéquat ou non adéquat selon le degré d'alphabétisation du public, selon qu'il est adressé de façon anonyme ou pas, etc. Une observation participante ou au contraire à l'insu de la personne observée sera pertinente ou non selon l'objet et les circonstances de l'étude.

Enfin, il ne suffit pas de disposer d'informations pertinentes et recueillies de façon adéquate, encore faut-il pouvoir montrer que la façon de les traiter permet d'approcher, quantitativement ou qualitativement selon les cas, les variables ou dimensions ou facteurs mentionnés dans le cadre problématique. Dans le cas de la mesure, réfléchir à la façon la plus adéquate de combiner et de pondérer plusieurs indicateurs (les questions du questionnaire ou les catégories d'une grille d'observation ou plusieurs indices statistiques) est une question de pertinence méthodologique. Dans une démarche qualitative, identifier parmi les nombreuses informations recueillies celles qui sont pertinentes, compte tenu du problème posé et du cadre problématique (établi au départ ou construit en cours de route), chercher à les réduire en les catégorisant, construire des profils de sens par mise en relation des catégories... sont autant de questions de pertinence.

2.1.3. *La pertinence de la communication*

Ne pas se tromper dans l'énonciation des conclusions de la recherche et dans la façon de les énoncer relève aussi de la pertinence. Il existe encore trop de rapports de recherches où n'apparaissent pas de véritables conclusions: le rapport s'arrête avec la présentation des résultats de recherche et la conclusion est absente ou n'est que de pure forme. Que faut-il nécessairement communiquer dans une conclusion? Comment le communiquer?

Puisqu'un problème a été posé, qu'un cadre conceptuel a été énoncé, qu'un cadre problématique a été émis traduisant un jeu d'hypothèses, que des informations ont été recueillies et traités pour obtenir des résultats présentés, la conclusion doit pouvoir apporter des réponses ou, du moins, des tentatives de réponses. Le problème de la pertinence des conclusions est de savoir si le chercheur s'est posé une série de questions fondamentales et si les conclusions y font allusion clairement. La première est sans doute de

pouvoir se demander si, au terme du processus, la question de recherche a été bien posée, reste pertinente telle qu'elle est posée ou s'il faut ajuster la façon de la poser. Dans le prolongement, la seconde est de se demander si les concepts identifiés sont et restent pertinents, si certains aspects de leur définition sont à revoir, si d'autres concepts n'apparaissent pas comme nécessaires. De la même façon, la troisième question est relative au cadre problématique énoncé: toutes les composantes de celui-ci sont-elles nécessaires? n'en faut-il pas d'autres? que penser des hypothèses relationnelles entre les composantes? n'en faut-il pas d'autres? dans ce cas, l'auteur a-t-il tenté de formuler des hypothèses alternatives argumentées sur la base des résultats? Finalement, la question centrale est bien de savoir si le chercheur a pensé à communiquer la réponse ou les réponses alternatives à donner à la question de départ ou à la question ajustée en cours de route.

La pertinence de la communication est aussi une question de forme. Le langage scientifique n'est pas un langage littéraire, ni administratif, ni journalistique. Les destinataires de la communication appartiennent à la communauté scientifique qui a pour fonction de légitimer la qualité des recherches conduites dans leur domaine de compétences. Toute conclusion doit pouvoir être l'objet d'une évaluation par les lecteurs du rapport de recherche. Le langage doit donc être clair, précis, argumenté, susceptible de permettre à l'évaluateur de formuler son propre avis.

2.2. La validité

Le concept de validité est certainement le concept dont les chercheurs et les méthodologues de la recherche parlent le plus, au point d'avoir distingué une infinité de formes spécifiques de la validité. En voici une liste non exhaustive: validité de construit, de construct, hypothético-déductive, structurelle, de contenu, de différenciation, curriculaire, discriminante, logique, définitoire, empirique, corrélationnelle, critérielle, factorielle, concourante, concomitante, apparente, de statut, congruente, convergente, prédictive, écologique, statistique... Si une telle liste existe, c'est que les méthodologues se sont penchés sur de nombreux cas d'espèce. Si ceci montre que la recherche est plurielle dans ses formes et ses visées, il ne faut pas oublier que la fonction commune de toute recherche est de produire de la connaissance scientifiquement fondée. C'est en regard de cette fonction commune et première que nous analysons le critère de validité. Rappelons que, fonda-

mentalement, se poser la question de la validité, c'est se demander: "Y a-t-il adéquation entre ce que je fais et ce que je déclare faire?". Ici à nouveau, les trois aspects essentiels de la démarche scientifique doivent faire l'objet de cette question: la conceptualisation, le recueil et le traitement des données, l'énonciation des conclusions.

2.2.1. La validité conceptuelle

Le chercheur peut avoir identifié les concepts pertinents pour étudier le problème posé, mais l'utilisation des concepts tout au long du processus de recherche et dans le rapport communiqué aux pairs peut révéler quelques failles de validité. C'est le cas chaque fois que les phénomènes dont parlent effectivement le chercheur ne correspondent pas aux concepts déclarés ou annoncés; et réciproquement. Prenons quelques exemples rencontrés fréquemment. Un chercheur pose le problème de la motivation en contexte scolaire, qu'il définit par ailleurs de façon adéquate, mais on s'aperçoit tout au long du rapport qu'il est plus question de la satisfaction des acteurs concernés que de la motivation. De nombreuses recherches sur l'efficacité déclarent évaluer des compétences et définissent la compétence comme le savoir agir qui consiste à mobiliser les ressources pertinentes pour effectuer des tâches complexes ou résoudre des situations problèmes, mais de fait le processus de recherche et le rapport ne font allusion qu'aux ressources à évaluer et non au pouvoir de mobilisation.

2.2.2. La validité méthodologique

Ici encore, le chercheur peut annoncer une méthodologie pertinente et dans les faits mettre en œuvre une méthodologie quelque peu différente. La question de la validité doit être posée sur tous les aspects de la méthodologie: le rapport entre l'échantillon et la population visée; le rapport entre le dispositif de recueil d'informations annoncé et celui réellement mis en œuvre; le rapport entre les informations (quantitatives ou qualitatives selon les cas) à récolter et celles qui sont effectivement récoltées (c'est-à-dire mesurées ou catégorisées selon les cas); le rapport entre le degré de validité déclarée par le chercheur et le degré réel de validité.

Le rapport entre l'échantillon réel et la population annoncée est un des premiers aspects à examiner, que la taille de l'échantillon soit importante ou qu'elle se résume à l'unité comme dans une étude de cas. Quoi qu'il

en soit, il s'agit toujours de donner au lecteur les informations essentielles pour contrôler ce qui est déclaré et voulu par le chercheur: ainsi, fournir des informations sur les caractéristiques des sujets est strictement nécessaire et encore trop souvent omis; une étude de cas qui annonce porter sur un sujet prototypique doit pouvoir le démontrer. Si, comme ces deux exemples le montrent, on parle le plus souvent de l'échantillon et de la population des sujets, n'oublions pas qu'il existe deux autres types d'échantillons et de populations trop souvent ignorés: l'échantillonnage et la population des informations susceptibles d'être recueillies; l'échantillonnage et la population des circonstances ou des contextes. Illustrons cela par deux exemples. Un test de performances est susceptible d'être composé d'une population quasi infinie d'items; finalement, quelle est la validité de l'échantillon des items retenus? Un récit de vie peut porter sur un nombre important d'épisodes; sur quels épisodes portent effectivement le recueil et le traitement de l'information?

Le rapport entre le dispositif annoncé de recueil d'informations et le dispositif effectivement mis en place peut concerner de nombreux aspects selon les cas? Citons quelques exemples. Le chercheur annonce trois entretiens par sujet (au début, au milieu et en fin de processus), mais on ne sait pas clairement si tous les entretiens ont été menés et, dans la négative, chez quels sujets, pour quelles raisons et quelles sont les implications. Les modalités d'application d'un test standardisé sont clairement explicitées, mais le rapport ne démontre pas que les modalités ont été respectées de fait. On annonce que tel groupe sera l'objet de la démarche expérimentale traduisant l'hypothèse du chercheur et que tel groupe sera un groupe témoin; il n'est pas impossible que, faute des vérifications nécessaires, le groupe témoin pratique davantage la démarche dite expérimentale que le groupe déclaré expérimental.

Le rapport entre l'information à récolter (fait, opinion, perception de soi, perception d'autrui, représentation sociale, conception, variable mesurée, catégorie... selon les cas) et celle effectivement récoltée est sans doute ce qui a été le plus étudié par les méthodologues férus de trouver des techniques de validation (ceci explique d'ailleurs les nombreuses appellations liées au concept de validité). Citons ici quelques exemples parmi d'autres. Dans un test, 10 items sont censés mesurer la compréhension de concepts étudiés et dix autres censés mesurer la capacité à inférer: cette volonté déclarée est-elle vérifiée par les indices alpha de Cronbach ou par les facteurs dégagés

d'une analyse factorielle? Des indices comme des tapes dans le dos ou des coups de points échangés par des adolescents observés dans une cours de récréation sont-ils des indices d'agressivité ou des indices d'amitié?

Enfin, rappelons que tout chercheur est censé indiquer dans son rapport le degré de validité qu'il peut raisonnablement déduire des différentes démarches méthodologiques effectives de sa recherche et de fournir les informations nécessaires pour que le lecteur puisse à son tour faire sa propre évaluation du degré de validité méthodologique de la recherche. Le chercheur lui-même a-t-il donné sa propre évaluation? Si oui, l'a-t-il argumentée? Son argumentation est-elle suffisante? N'aurait-il pas fallu fournir d'autres informations pour apprécier le degré de validité réelle? Le degré de validité est-il suffisant, compte tenu du fait qu'il n'est jamais possible d'obtenir une validité méthodologique à 100%?

2.2.3. La validité interne et externe des conclusions

La validité des conclusions est le degré d'adéquation des conclusions énoncées dans le rapport de recherche et des conclusions que l'on peut effectivement tirer. On parlera de validité interne lorsque les conclusions sont examinées et évaluées sur le plan logique; on parlera de validité externe lorsque les conclusions sont examinées sur le plan de la généralisabilité. Dans les deux cas, il s'agit d'évaluer la qualité des inférences effectuées dans les conclusions.

Le degré de validité interne des conclusions dépend étroitement: (1) du degré de validité conceptuelle et du degré de validité méthodologique; (2) de l'inférence permise compte tenu de ces degrés de validité. Encore faut-il qu'il y ait de réelles conclusions, c'est-à-dire de conclusions en rapport avec le problème posé au départ (question de pertinence). Sur le plan interne, une conclusion ne sera pas valide, par exemple, si l'on affirme qu'est vérifiée l'hypothèse de la supériorité de tel type de pratiques pédagogiques, alors qu'on est loin d'être sûr que les pratiques réellement mises en œuvre obéissent aux caractéristiques déclarées. Telle autre conclusion, affirmant que le niveau socio-économique est un déterminant important, ne pourra être déclarée valide si la toute grande majorité des sujets observés appartiennent à la même catégorie socioéconomique.

Le degré de validité externe des conclusions concerne le degré de généralisabilité des conclusions. La littérature méthodologique décrit toute une série de biais qui invalident la généralisabilité des conclusions à des populations plus larges de sujets, d'informations ou de contextes. Ainsi, par exemple, le fait d'avoir subi un test dans l'expérience a pu orienter l'attention des sujets vers les résultats escomptés par le chercheur et on ne pourra généraliser les conclusions que si dans les pratiques ultérieures un test fera partie du dispositif. Le fait que le dispositif de recherche a utilisé des pratiques ou des supports inhabituels est peut-être la cause principale des performances obtenues. La recherche a été effectuée avec des sujets volontaires et les conclusions ne seront vraisemblablement généralisables que pour un tel public. Ces exemples, parmi bien d'autres, induisent l'idée de l'importance de la conscience des justes limites de sa recherche (ni trop sévère, ni trop indulgent; mais une grande lucidité).

2.3. La fiabilité

Disons d'emblée que nous préférons le concept de fiabilité à celui de fidélité ("reliability"). Ce dernier est trop restrictif car il est habituellement défini comme l'écart entre les scores observés et les scores dits "vrais". Non seulement, on parle de "scores", ce qui n'implique que certains types de recherches, mais on parle de scores "vrais", ce qui nous paraît un concept très difficile à préciser et peut amener à confondre fidélité et certaines formes de validité. Le concept de fiabilité a une acceptation plus large et convient à tout type de recherche, car il signifie que *«ce qui est affirmé dans la recherche est indépendant de l'auteur de la recherche et pourrait donc être affirmé par d'autres chercheurs»*. Comme pour la pertinence et la validité, la fiabilité peut être examinée sous trois aspects: la conceptualisation, le recueil et le traitement des données, les conclusions.

2.3.1. La fiabilité des concepts

Les concepts ne sont pas fiables lorsque leur utilisation est trop dépendante de l'auteur de la recherche. C'est le cas lorsque de nouveaux concepts sont apportés par l'auteur, alors qu'il existe déjà des concepts reconnus par la communauté scientifique et qui traduisent aussi bien les phénomènes décrits. C'est encore le cas lorsque, de façon justifiée, un nouveau concept est créé, mais la définition utilise des termes qui risquent d'être compris

différemment par des chercheurs d'horizons différents. C'est encore le cas lorsque des chercheurs actuels utilisent encore des concepts dans leur acception historique originale, alors que des chercheurs contemporains ont montré, tout en reconnaissant les avancées que les chercheurs antérieurs ont amenées, la confusion qu'ils ont engendrée avec le développement de la recherche. Un exemple frappant est le concept d'évaluation sommative qui historiquement a permis de faire émerger le concept d'évaluation formative (apport de l'Américain Scriven). Or les recherches ultérieures dans le domaine de l'évaluation (voir par exemple De Ketele, 1993, 2006) ont montré que cette distinction conduisait à confondre la fonction et la démarche de l'évaluation: l'évaluation sommative est de l'ordre de la démarche (faire une somme pour obtenir une note ou un score), tandis que l'évaluation formative est de l'ordre de la fonction (viser l'amélioration de l'apprentissage en cours). Beaucoup de chercheurs actuels utilisent le concept d'évaluation sommative, alors qu'en fait ils désignent la fonction d'évaluation certificative, laquelle peut être l'objet d'une démarche sommative (on certifie par une note) ou d'une démarche descriptive (comme dans le bulletin descriptif ou dans un certificat) ou d'une démarche herméneutique (on certifie en donnant du sens à un ensemble d'indices de différentes natures, quantitatives ou qualitatives).

2.3.2. *La fiabilité des démarches méthodologiques*

Les démarches méthodologiques (identification des unités d'échantillonnage, application du dispositif de recueil d'informations, modalités de traitement de l'information...) ne sont pas fiables lorsqu'elles dépendent de l'auteur qui la met en œuvre. Quand il s'agit de démarches quantitatives, on retrouve ici les nombreux coefficients dits de fidélité ("reliability") qui ont été créés par les méthodologues de la mesure et on fait aussi référence aux travaux des docimologues (par exemple: Piéron, 1973; Noizet et Caverni, 1978) lorsqu'il s'agit de mesures de performances. Quand il s'agit de démarches qualitatives, la fiabilité s'exprimera surtout en termes d'accord ou de triangulation. Par exemple, s'il s'agit de démarches d'observation (ou d'évaluation ou d'analyse de corpus), il importe d'indiquer les résultats des vérifications qui s'imposent pour vérifier la fiabilité des observateurs (ou des évaluateurs ou des analystes). Selon les cas, cela peut être l'un ou l'autre des accords suivants: accord entre des observateurs différents; accord chez un même observateur à travers la durée; accord entre plusieurs équipes d'observateurs; accord à l'intérieur d'une même équipe d'observateurs; accord

entre un observateur et un protocole. Ce qui vient d'être dit en rapport avec l'observation est évidemment transposable pour des démarches d'évaluation ou d'analyse de corpus. Dans d'autres situations de recherche, on utilisera les principes de la triangulation. De nombreux cas de figures existent qui n'ont évidemment pas la même valeur en termes de fiabilité: une même information obtenue trois fois par la même personne; par trois personnes différentes de même statut; par trois personnes de statut différent; dans trois contextes différents; provenant de trois sources méthodologiquement différentes (exemple: questionnaire, entretien, observation)... D'autres combinaisons existent. Notre expérience de rédacteur et de membres de comités scientifiques de revues nous amène à dire que trop de rapports de recherches utilisant des démarches qualitatives ne mentionnent pas les démarches de vérification de leurs démarches méthodologiques. Par ailleurs, cette même expérience nous montre que, si de nombreux rapports de recherches utilisant des démarches quantitatives fournissent des coefficients de fiabilité (de fidélité), ils oublient souvent de communiquer comment la quantification a été élaborée. Déjà en 1976, dans son ouvrage *Méthodes en Sciences sociales* couronné par l'Académie des Sciences Morales et Politiques, Madeleine Grawitz reprochait à certains chercheurs quantitativistes de se réfugier derrière des démarches techniques très sophistiquées pour vérifier la fiabilité de leurs mesures, mais oubliaient qu'une bonne mesure est fondée au départ sur une bonne qualification du phénomène à évaluer.

2.3.3. *La fiabilité des conclusions*

Le problème de la fiabilité des conclusions est d'examiner si les conclusions énoncées dans un rapport de recherche sont suffisamment indépendantes du chercheur au point de pouvoir être énoncées de façon équivalente par d'autres chercheurs. L'exemple le plus typique est sans doute le cas de la conclusion de l'affirmation d'un effet (tel facteur est à la base de tel effet) sur la base de résultats positifs mais non statistiquement significatifs, alors que la prudence du scientifique est de se dire que l'effet positif obtenu peut être dû à de nombreux facteurs conjoints (erreur de mesure, biais dans la mise en place du dispositif, covariances, variables nichées dans l'échantillonnage, etc.). Le manque de fiabilité des conclusions est aussi souvent dû à la formulation des conclusions qui peuvent être comprises de différentes façons selon les lecteurs du rapport.

3. De quelques questions actuelles

Parmi les nombreuses questions que l'on peut se poser sur la qualité de la recherche, nous voudrions en évoquer trois très rapidement: les relations entre pertinence, validité et fiabilité; les tendances à hiérarchiser les différents types de recherche; la question de la pertinence sociale.

Les trois qualités que nous avons développées dans la section précédente forment un tout cohérent comme le montre le tableau suivant:

Les critères incontournables

	PERTINENCE	VALIDITÉ	FIABILITÉ
CONCEPTUALISATION	Pertinence théorique	Validité conceptuelle	Fiabilité des concepts
RECUEIL ET TRAITEMENT DES DONNÉES	Pertinence méthodologique	Validité méthodologique	Fiabilité des démarches méthodologiques
ÉNONCIATION DES CONCLUSIONS	Pertinence de la communication	Validité interne et externe (généralisabilité)	Fiabilité des conclusions

Figure 3

Ces critères sont incontournables pour toute recherche scientifique en éducation, au sens défini précédemment. Il va de soi que ces critères et les aspects sur lesquels ils portent ne sont pas indépendants. L'idéal serait qu'une recherche soit à 100% pertinente, valide et fiable. Cet idéal n'existe pas, même dans les sciences dites dures (cfr. les travaux des épistémologues sur l'équation personnelle, par exemple); à plus forte raison dans les sciences humaines qui sont des sciences du complexe. On se trouve donc toujours dans des situations intermédiaires avec des degrés divers de pertinence, de validité et de fiabilité. Mais nous sommes persuadé que le critère de pertinence est le critère premier, même s'il est le critère le moins technique dans

le sens où il n'existe pas de modèle mathématique pour la vérifier, contrairement à la fiabilité et à certains aspects de la validité où l'arsenal mathématique est très large. A quoi servirait une recherche dont les indices de validité et de fiabilité sont excellents si elle manque de pertinence! Cette situation est loin d'être rare tant les chercheurs sollicités par les revues scientifiques sont souvent d'excellents techniciens de la méthodologie de la recherche et sont particulièrement sensibles aux indices mathématiques liés à la validité et à la fiabilité. Un regard rétrospectif sur l'histoire des sciences montre que les apports les plus importants à la production de connaissances scientifiques étaient le fait de recherches loin d'être parfaites sur les plans techniques, mais où la façon de poser le problème, le dispositif mis en place et les conclusions qui en découlaient étaient d'une grande pertinence. Bien plus, les rapports de ces recherches n'étaient pas nécessairement publiés dans des revues que nous classerions aujourd'hui au niveau A et les auteurs de ces recherches auraient souvent eu, s'il avait existé à l'époque, un "citation index" médiocre. Ne jetons cependant pas le bébé avec l'eau du bain, car à quoi servirait une recherche dont les indices de fiabilité seraient désastreux, car on n'aurait aucune assurance sur la validité et la pertinence.

L'esprit humain est ainsi fait qu'il a besoin de hiérarchiser les produits et les personnes. Malgré la très grande modestie qu'implique l'esprit scientifique, le monde des chercheurs n'échappe pas à cette tendance et on assiste à des jugements de valeur sur les différents paradigmes de recherche et une lutte de pouvoir entre les partisans de ces paradigmes dans un sens ou dans l'autre. Malgré qu'à cette époque, le slogan "publish or perish" n'était pas encore aussi impératif que maintenant, Madeleine Grawitz (1976) disait déjà qu'il n'existe pas des types de recherches qui seraient plus nobles et d'autres moins nobles, que la lutte entre les quantitativistes et les qualitativistes est absurde, puisque l'essentiel est de produire de la connaissance nouvelle. Ainsi disait-elle qu'«un grand déploiement de procédés mathématiques cachait mal la pauvreté des résultats, le manque de valeur et même de rigueur des éléments qualitatifs, hâtivement recueillis et peu significatifs» (*Ibidem*, 361). Elle avait également la dent dure contre les chercheurs qui se réfugiaient dans le qualitatif faute de pouvoir faire du quantitatif. Or, disait-elle, les démarches qualitatives et quantitatives se complètent: la qualification appelle la quantification (comme compter, estimer l'importance...) et la quantification suppose une bonne qualification. Le chercheur ne peut conduire une recherche de qualité sans avoir reçu une formation solide sur les deux types de démarches.

Un autre aspect qui fait débat actuellement est relatif au critère de pertinence sociale dont certains chercheurs voudraient faire un critère incontournable pour toute recherche dans des domaines comme ceux de l'éducation et de la formation. Dans une communication récente devant une assemblée internationale de chercheurs (Louvain-la-Neuve, 16 novembre 2009), Léopold Paquay et moi-même défendions l'idée que si la pertinence scientifique était un critère incontournable de toute recherche scientifique (puisque toute recherche a pour fonction de produire des connaissances scientifiques), on ne pouvait dire la même chose de la pertinence sociale (le fait que la recherche est conçue essentiellement pour servir la pratique). S'il est vrai que des recherches ont été conçues dès le départ dans cette visée et ont pu produire des connaissances scientifiques nouvelles, il en est bien d'autres qui ont été conçues sans aucune visée de cet ordre et ont produit des connaissances scientifiques dont la pertinence sociale n'est apparue que quelques décennies plus tard. Dans le même ordre d'idée, nous défendions l'idée que la recherche en éducation ne peut avoir une fonction normative ou prescriptive, c'est-à-dire prescrire des solutions "clés sur porte" ou des principes à appliquer. Les recherches en éducation produisent des connaissances nouvelles dans des contextes de recherche et avec les outils de la recherche; aux praticiens, de faire le travail d'ingénierie nécessaire pour s'inspirer des résultats de la recherche, de les adapter aux contextes qui sont les leurs, de se servir des théories et des modèles établis par les chercheurs comme des grilles d'analyse et non comme des théories et des modèles à appliquer.

Présentation de l'Auteur: Jean-Marie De Ketele est professeur émérite à l'Université Catholique de Louvain (UCL) Belgique; membre fondateur du GIRSEF (Groupe Interfacultaire de Recherche sur les Systèmes d'Éducation et de Formation); titulaire de la CUSE (Chaire UNESCO en Sciences de l'Éducation); directeur de la collection Pédagogies en développement – De Boeck et rédacteur européen de la revue *Mesure et évaluation en éducation*.

Bibliografia

- BANDURA, A. (2003), *Auto-efficacité. Le sentiment d'efficacité personnelle*, Bruxelles, De Boeck.
- CADEI, L. (2006), «L'incertezza delle cifre: i fatti e gli artefatti», in L. PATI (a cura di), *Infanzia violata e solidarietà sociale. Elementi per un'interpretazione pedagogica*, Milano, I.S.U., 111-122.

- CAMPBELL, D.T. , STANLEY, J. (1966), *Experimental and quasi-experimental designs for research*, Chicago, Rand Mac Nally.
- DE KETELE, J.-M. (2006), «Contrôles, examens et évaluation», in J. BEILLEROT, N. MOSCONI (dir.), *Traité des sciences et pratiques de l'éducation*, Paris, Dunod.
- DE KETELE, J.-M., MAROY, CH. (2006), «Quels critères de qualité pour les recherches en éducation», in L. PAQUAY, M. CRAHAY, J.-M. DE KETELE, *L'analyse qualitative en éducation. Des pratiques de recherche aux critères de qualité*, Bruxelles, De Boeck.
- DE KETELE, J.-M., ROEGIERS, X. (1993), *Méthodologie du recueil d'informations*, Bruxelles, De Boeck.
- GRAWITZ, M. (1976), *Méthodes des sciences sociales*, Paris, Dalloz.
- NOIZET, G., CAVERINI, J.-P. (1978), *Psychologie de l'évaluation scolaire*, Paris, PUF.
- PAQUAY, L., CRAHAY, M., DE KETELE, J.-M. (2006), *L'analyse qualitative en éducation. Des pratiques de recherche aux critères de qualité*, Bruxelles, De Boeck.
- PIÉRON, H. (1973), *Examens et docimologie*, Paris, PUF.
- VIAU, R. (2004), *La motivation en contexte scolaire*, Bruxelles, De Boeck.