



HAL
open science

Innovative and alternative research methods in economics and business administration

Anne Bartel-Radic

► **To cite this version:**

Anne Bartel-Radic (Dir.). Innovative and alternative research methods in economics and business administration. EIKV-Schriftenreihe zum Wissens- und Wertemanagement, 33, 2019, EIKV-Schriftenreihe zum Wissens- und Wertemanagement. hal-03566071

HAL Id: hal-03566071

<https://hal.univ-grenoble-alpes.fr/hal-03566071>

Submitted on 11 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bartel-Radic, Anne (Ed.)

Research Report

Méthodes de recherche innovantes et alternatives en économie et gestion

EIKV-Schriftenreihe zum Wissens- und Wertemanagement, No. 33

Provided in Cooperation with:

European Institute for Knowledge & Value Management (EIKV),
Luxemburg

Suggested Citation: Bartel-Radic, Anne (Ed.) (2019) : Méthodes de recherche innovantes et alternatives en économie et gestion, EIKV-Schriftenreihe zum Wissens- und Wertemanagement, No. 33, European Institute for Knowledge & Value Management (EIKV), Rameldange

This Version is available at:

<http://hdl.handle.net/10419/206541>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

EIKV-Schriftenreihe zum
Wissens- und Wertemanagement

Méthodes de recherche innovantes et alternatives en
économie et gestion

--

Innovative and alternative research methods in
economics and business administration

Coordonné par / Coordinated by
Anne Bartel-Radic

IMPRESSUM

EIKV-Schriftenreihe zum Wissens- und Wertemanagement

Herausgeber: André Reuter, Heiko Hansjosten, Thomas Gergen

© EIKV Luxemburg, 2016, 2017, 2018, 2019

European Institute for Knowledge & Value Management (EIKV)

c/o M. André REUTER – 8, rue de la Source

L-6998 Hostert - GD de Luxembourg

info@eikv.org

www.eikv.org

Sommaire / Table of Contents

PRÉFACE	2
FOREWORD	4
<hr/>	
QUANTIFIER DES ÉTUDES DE CAS QUALITATIVES AVEC QCA : ENJEUX ET PROCESSUS DE LA MÉTHODE À PARTIR D'UNE RECHERCHE SUR LA COLLABORATION INTER-ORGANISATIONNELLE	6
QUANTIFYING QUALITATIVE CASE STUDY RESEARCH WITH QCA: IMPLICATIONS AND PROCESS OF THE METHOD FROM THE EXAMPLE OF A STUDY ON INTER-ORGANIZATIONAL COLLABORATION	26
<hr/>	
LES MÉTHODES EXPÉRIMENTALES DANS LA RECHERCHE EN MANAGEMENT INTERNATIONAL	44
EXPERIMENTAL METHODS IN INTERNATIONAL MANAGEMENT RESEARCH	59
<hr/>	
APPORTS DE LA THÉORIE MODERNE DE LA MESURE (IRT) À L'ÉTUDE DES ECHELLES D'INTERVALLE	74
IRT APPROACH FOR RATING SCALES: APPLICATIONS FOR NORMAL AND NON-NORMAL DISTRIBUTIONS	95
<hr/>	
L' APPORT D'UN LOGICIEL CAQDAS À L'ANALYSE DES RÉCITS DE VIE : L'EXEMPLE DE NVIVO	119
A METHOD FOR ANALYZING LIFE HISTORIES USING CAQDAS SOFTWARE	135
<hr/>	
MOBILISER L'EMPATHIE DANS NOS MÉTHODES DE RECHERCHE : DES PROPOSITIONS INSPIRÉES DU DESIGN THINKING	149
MOBILIZING EMPATHY IN OUR RESEARCH METHODS: PROPOSITIONS INSPIRED BY DESIGN THINKING	159
<hr/>	
FAIRE DE LA RECHERCHE-ACTION DANS UNE ÉQUIPE PROJET INTERNATIONALE	168
DOING ACTION RESEARCH IN A GLOBAL PROJECT TEAM	182

Préface

Cet ouvrage regroupe des contributions sur des méthodes de recherche pouvant être qualifiées d'innovantes ou d'alternatives en Sciences économiques et de gestion. L'idée est née lors de discussion avec des collègues de ces disciplines. Nous avons chacun.e exploré puis approfondi des méthodologies plus ou moins atypiques dans nos recherches respectives. Des méthodologies atypiques par leur nouveauté, mais une nouveauté toute relative : souvent, ces méthodologies sont déjà plus couramment utilisées dans une autre discipline, mais encore méconnues en économie ou en gestion. C'est dans le souhait de partager ces approches que nous avons décidé de les réunir dans ce livre, qui contient six contributions différentes. Certains de nos textes étaient en français, d'autres en anglais. Plutôt que de tout traduire en une seule langue et de « priver » une partie de nos lecteurs potentiels de l'accès à nos textes, nous avons fait le choix d'un livre bilingue. Ainsi, tous les chapitres sont successivement inclus dans les deux versions, française et anglaise.

Les méthodologies présentées dans ce livre sont très variées : des méthodologies intrinsèquement qualitatives comme la recherche-action et l'analyse biographique, des méthodologies quantitatives comme l'expérimentation et l'analyse des échelles d'intervalle, ainsi qu'une méthodologie à mi-chemin entre « quali » et « quanti », la méthode QCA. Ce premier chapitre sur l'analyse qualitative comparative (*qualitative comparative analysis*, QCA) est écrit par Amélie Artis, Anne Bartel-Radic et Hezam Haidar. Nous discutons la quantification d'études de cas qualitatives avec la méthode QCA à l'exemple d'une recherche sur la collaboration inter-organisationnelle. L'étude explore, à l'aide de données secondaires, les collaborations entre des entreprises et des organisations à but non lucratif visant à implémenter des projets dans les énergies renouvelables. Une analyse QCA permet d'identifier les facteurs de réussite de ces collaborations. Le chapitre détaille les opportunités et limites que représente l'utilisation de cette méthode, et ses implications pour les chercheurs utilisant des études de cas qualitatives.

Le deuxième chapitre, par Anne Bartel-Radic, Philippe Mouillot et Danielle Taylor, porte sur les méthodes expérimentales dans la recherche en management international. Les méthodes expérimentales sont depuis longtemps employées en Sciences naturelles afin de tester une hypothèse en contrôlant les conditions d'expérimentation. Ces méthodes ont été reprises en Sciences économiques et de gestion ces dernières décennies, et sont devenues une référence dans certaines sous-disciplines comme le marketing. Or, les expérimentations restent rares dans la recherche en management international. Dans ce chapitre, nous présentons les difficultés, mais aussi les avantages des méthodologies expérimentales dans cette sous-discipline à travers l'exemple du projet InterCCom, portant sur la compétence interculturelle. Grâce au développement d'un jeu sérieux informatisé dans le cadre de ce projet, les comportements des membres d'une équipe internationale virtuelle sont mesurés, ce qui permettra de répondre à des problématiques de recherche variées.

La troisième contribution, par Ping Lei, Maud Damperat, Florence Jeannot et Alain Jolibert, porte sur un problème auquel sont confrontés les chercheurs mobilisant des méthodes quantitatives, et notamment les échelles d'intervalle : l'absence de normalité. Ce chapitre présente l'intérêt de l'IRT (Item Response Theory), et plus particulièrement du modèle IRT de

crédit partiel, adapté aux échelles d'intervalle. Il concerne la construction d'échelles en l'absence de normalité des variables de mesure, et détaille les apports de l'IRT aux tests issus de la théorie classique de la mesure. Sur la base d'une collecte de données portant sur une échelle asymétrique de satisfaction, ce chapitre décrit les étapes d'une analyse IRT de crédit partiel et les indicateurs utiles (*i.e. paramètres de discrimination et de difficulté*) pour la sélection d'items pertinents. L'IRT s'avère intéressante en complément des tests classiques en proposant un complément d'information qui pallie certains de leurs écueils.

Les trois derniers chapitres portent sur des méthodes qualitatives. Dans le chapitre 4, Amélie Artis et Anaïs Bovet présentent une méthode d'analyse des biographies de vie à l'aide d'un logiciel CAQDAS. Les récits de vie sont une méthodologie fortement ancrée dans la sociologie empirique, mais cette méthode offre des perspectives intéressantes pour l'économie et la gestion car elle permet de mieux comprendre les interactions entre l'individu, le groupe et la société par la façon dont l'individu se raconte. Les nombreux progrès réalisés par les logiciels d'analyse de données textuelles pour traiter des quantités de données importantes facilitent l'utilisation des récits de vie. Ce chapitre présente les fondements de la méthode des récits de vie, puis expose son traitement avec le logiciel d'analyse de données textuelles NVivo.

Dans le chapitre 5, Gaëlle Dechamp apporte son regard sur la recherche-action dans le domaine du design, et plaide pour l'empathie du chercheur. En Sciences de gestion, les émotions ont connu un intérêt grandissant à partir du milieu des années 90. Le rôle des événements organisationnels comme « amorce émotionnelle » a ainsi été théorisé dans l'*affective events theory*. Le chercheur qui souhaite comprendre des comportements dans et de l'entreprise peut prendre en compte cette dimension émotionnelle en étant capable d'empathie pour s'approprier et comprendre en profondeur ce qui se joue dans l'organisation. Les outils de la méthodologie du Design Thinking (méthode d'innovation centrée sur l'humain, mobilisant pensée analytique et pensée intuitive) sont particulièrement appropriés pour mobiliser l'empathie dans la recherche en Gestion.

Enfin, le chapitre 6 propose une seconde contribution sur la recherche-action. Ce chapitre se base sur l'expérience de Fabienne Münch qui, en tant que cheffe d'une équipe de design et développement internationale, a mis en oeuvre une démarche de recherche-action en deux temps. Durant le travail de l'équipe, elle a adopté une démarche de « réflexion en action » permettant de prendre du recul sur les situations. Puis, après la fin du projet, une collaboration avec Anne Bartel-Radic, une chercheuse n'ayant pas connu l'équipe, a donné lieu à davantage de théorisation de l'expérience vécue. Ce chapitre passe en revue la littérature méthodologique pour montrer en quoi la recherche-action est une méthode intéressante en management international, détaille deux situations vécues, et propose des recommandations aux chercheurs souhaitant mener une recherche-action en contexte interculturel.

Plusieurs auteurs ayant contribué à ce livre (Anaïs Bovet, Hezam Haidar, Fabienne Münch et Danielle Taylor) sont lors de sa publication au début, au milieu ou vers la fin d'une thèse de doctorat. Leurs contributions témoignent de leur niveau d'expertise sur ces méthodologies, et j'espère que cette publication leur sera utile pour la suite de leurs parcours de chercheurs !

A Grenoble, en octobre 2019
Anne Bartel-Radic

Foreword

This book brings together various contributions on research methods that can be qualified as innovative or alternative in the fields of economics and management. The idea came about during a discussion with colleagues from these disciplines. We had all explored and deepened our understandings of methodologies that are more or less atypical in our respective research. These are methodologies that are atypical as defined by their novelty, but even so this is quite relative; often, these methodologies are already commonly used in other disciplines, but still unknown in economics or management. We all wanted to share these approaches, but they could not easily find their place in other publication formats; they were considered as not sufficiently innovative for academic journals centered on methodology and too varied to be able to constitute a complete collection. So, we decided to bring our research together in this book, which contains six different contributions. Some of our texts were in French, others in English. Rather than translate everything into one language and “deny” some of our potential readers access to our texts, we chose to write a bilingual book. Thus, all the six chapters are included in both French and English.

The methodologies presented in this book cover the entire spectrum including intrinsically qualitative methodologies such as action research and biographical analysis, quantitative methodologies such as experimentation and analysis of interval scales, as well as a methodology considered as halfway between “quali” and “quanti”: the QCA method. This first chapter on qualitative comparative analysis (QCA) is written by Amélie Artis, Anne Bartel-Radic and Hezam Haidar. We discuss the quantification of qualitative case study research with the QCA method on the example of a study of inter-organizational collaboration. The study explores collaborations between companies and not-for-profit organizations implementing projects in the renewable energy sector, using secondary data. Qualitative comparative analysis permits to identify the success factors of these collaborations. The chapter details the opportunities and limits of this method of analysis, and its implications for researchers using qualitative case studies.

The second chapter is written by Anne Bartel-Radic, Philippe Mouillot and Danielle Taylor and deals with experimental methods in international management research. Experimental methods have long been used in the natural sciences to test a hypothesis by controlling experimental conditions. These methods have also been used in economics and management in recent decades, and have become a reference in some sub-disciplines such as marketing. However, experiments are rare in international management research. In this chapter, we present the difficulties, but also the advantages of the experimental methodologies in this sub-discipline through the example of the InterCCom project, which deals with intercultural competence. Thanks to the development of a computerized serious game for this project, the behaviors of the members of a virtual, international team are measured. Various research questions are outlined.

The third contribution, written by Ping Lei, Maud Damperat, Florence Jeannot and Alain Jolibert, deals with a problem that researchers face using quantitative methods, including

interval scales: the lack of normality. This chapter presents the interest of the IRT (Item Response Theory), and more particularly, the IRT model of partial credit adapted to the interval scales. It concerns the construction of scales in the absence of normality of the variables of measurement and details the contributions of the IRT to the tests resulting from the classical theory of the measurement. Based on a collection of data on an asymmetric satisfaction scale, this chapter describes the steps of an IRT partial credit analysis and the meaningful indicators (i.e. discrimination and difficulty parameters) for the selection of relevant items. In addition to conventional tests, the IRT is interesting because it offers additional information that overcomes some common pitfalls.

The last three chapters deal with qualitative methods. In chapter four, Amélie Artis and Anaïs Bovet present a method for analyzing life histories using CAQDAS software. Life narratives is a methodology strongly rooted in empirical sociology, but this method offers interesting perspectives for other disciplines as it allows for a better understanding of the interactions between the individual, group and society through analyzing the way in which the history is told. In addition, the many advancements made by textual data analysis software in order to process large amounts of data invite researchers to use life stories. In this chapter, foundations of the method of life stories is first presented, then secondly its treatment with NVivo textual data analysis software is considered.

In the fifth chapter, Gaëlle Dechamp takes a look at action research in the field of design, and pleads for the researcher's empathy. In Management Research, emotions have been of growing interest since the mid-1990s. The role of organizational events as an "emotional trigger" has been theorized in affective events theory. The researcher who wishes to understand phenomena and behaviours in and of the company must take this emotional dimension into account and therefore be able to empathize in order to appropriate and understand in depth what is at stake in the organization. The tools of the Design Thinking methodology (a human-centred innovation method, which aims to synthesize analytical and intuitive thinking) are very appropriate to mobilize empathy in Management research.

Finally, chapter six presents the case of a global design and development project team in which the team leader, Fabienne Münch, has implemented two phases of action-research. During the teamwork, she implemented "reflection in action" that allowed to take a step back on situations. Later, after the end of the project, a collaboration with the second author, Anne Bartel-Radic, led to further theorization. This chapter reviews the methodological literature to show that action research is an interesting method for international management research and presents how action research was conducted during two events that occurred during the project teamwork. Finally, recommendations are made to researchers who intend to implement action research in intercultural teams.

Several authors who have contributed to this book (Anaïs Bovet, Hezam Haidar, Fabienne Münch and Danielle Taylor) are, at the time of publication, at the beginning, middle or towards the end of a doctoral thesis. Their contributions attest to their level of expertise on these methodologies, and I hope that this publication will be helpful for their careers in research!

Grenoble, October 2019
Anne Bartel-Radic

Quantifier des études de cas qualitatives avec QCA : enjeux et processus de la méthode à partir d'une recherche sur la collaboration inter-organisationnelle

RÉSUMÉ

Ce chapitre discute la quantification d'études de cas qualitatives avec la méthode QCA à l'exemple d'une recherche sur la collaboration inter-organisationnelle. L'étude explore, à l'aide de données secondaires, les collaborations entre des entreprises et des organisations à but non lucratif visant à implémenter des projets dans les énergies renouvelables. Une analyse QCA permet d'identifier les facteurs de réussite de ces collaborations. Le chapitre détaille les opportunités et limites que représente l'utilisation de cette méthode, et ses implications pour les chercheurs utilisant des études de cas qualitatives.

MOTS CLÉS : analyse qualitative comparée (QCA), collaboration inter-organisationnelle, organisation sans but lucratif, secteur des énergies renouvelables

INTRODUCTION

Depuis plus de trente ans persiste en Sciences de gestion une opposition entre méthodologies de recherche quantitatives et qualitatives. Au-delà du fait que les deux catégories de méthodes répondent à des objectifs différents et complémentaires (Bergadaà et Nyeck, 1992), la quantification de données qualitatives peut contribuer à combler le « fossé » qui continue à les séparer. Une des méthodes de quantification des données qualitatives est la méthode QCA, « qualitative comparative analysis », aussi appelée analyse quali-quantitative comparée (Ragin, 1987 ; Rihoux, Marx et Alamos-Concha, 2014). Bien que son utilisation croissante en gestion a été prédite il y a plus de dix ans (Chanson et al., 2005), et malgré un réel essor depuis 2005 environ (Rihoux, Alamos-Concha, Bol, Marx et Rezsöhazi, 2013 ; Seny Kan, Adegbite, El Omari et Abdellatif, 2016), les chercheurs méconnaissent encore assez largement cette méthode. Son intérêt réside dans l'analyse de la complexité causale qui implique différentes combinaisons de conditions, capables de générer le même résultat sur des échantillons de taille intermédiaire.

L'objectif de ce chapitre n'est pas de décrire une nouvelle fois la méthode QCA et ses utilisations passées ou possibles pour l'étude des organisations, mais de discuter l'intérêt cette méthode parmi les différentes méthodes de quantification des données qualitatives dans un but didactique. Cette méthode initiée dans les Sciences politiques comparées est de plus en plus utilisée dans la sociologie et les Sciences de gestion, en particulier en stratégie (Wagemann et al., 2016, Rihoux et al., 2013) et en management (Seny Kan et al., 2016). L'enjeu pour son fondateur, C. C. Ragin (1987) était de développer une méthode en Sciences

sociales qui articule les approches « centrées sur les cas » et « centrées sur les variables » (Rihoux et al., 2014, p. 77). Même si d'autres utilisations sont possibles, et que des variantes sophistiquées et informatisées ont été développées, nous argumentons que la « version basique » de la QCA mérite d'être mieux connue par les chercheurs collectant des données qualitatives au sein d'études de cas.

Nous nous appuyons sur une recherche que nous avons menée sur les collaborations inter-organisationnelles, dans laquelle nous avons mobilisé la méthode QCA pour l'analyse des données collectées. C'est à lumière de cette étude que nous discuterons les opportunités et limites liées à cette méthode pour les chercheurs en Sciences sociales. A travers notre exemple, l'enjeu est de présenter le protocole d'une recherche afin de discuter les choix méthodologiques auxquels sont confrontés les chercheurs en Sciences sociales et les apports de la méthode QCA par rapport à d'autres méthodes. L'originalité est donc d'explicitier les « impensés » méthodologiques et leurs techniques « naïves » (Sims, 2010).

Dans notre exemple, nous nous sommes intéressés aux facteurs de succès de collaborations entre entreprises et organisations à but non lucratif (associations, ONG, organismes de recherche et publics...) visant à implémenter un projet dans le secteur des énergies renouvelables – dans la grande majorité des cas, il s'agit de constructions de centrales électriques utilisant des énergies renouvelables (solaire, hydraulique, éolien et géothermique).

Nous procéderons de la manière suivante : dans une première partie, nous présenterons le cadre théorique et la problématique de l'étude sur les collaborations entre organisations de natures différentes, et questionnerons l'opportunité de les aborder avec la méthode QCA par rapport à d'autres méthodologies. Dans la deuxième partie, nous détaillerons les données empiriques collectées sur trente cas de collaborations, et leur analyse avec la méthode QCA. Enfin, en partant de l'exemple de cette recherche, nous discuterons des apports et limites de la méthode QCA pour les Sciences de gestion et donnerons des recommandations quant à son utilisation.

1. CONTEXTE DE LA RECHERCHE : THÉMATIQUE ET CHOIX DE LA MÉTHODE D'ANALYSE

Nous présenterons d'abord brièvement la littérature sur le fond du sujet étudié, les collaborations inter-organisationnelles dans le secteur des énergies renouvelables, avant d'aborder l'intérêt de retenir la QCA parmi les différentes méthodologies envisageables afin de réaliser une étude sur le sujet.

1.1. Revue de la littérature sur les collaborations inter-organisationnelles dans les projets d'énergies renouvelables

Les énergies renouvelables sont un secteur porteur d'innovations technologiques mises en œuvre par de grandes entreprises multinationales. Souvent créées par des gouvernements, les grandes entreprises du secteur de l'énergie ont aujourd'hui de plus en plus d'injonctions à répondre aux enjeux de la transition énergétique (Reverdy, 2014). Pourtant ces sociétés

commerciales ne sont pas encore significativement engagées dans la transition, elles sont peu nombreuses à prendre des risques financiers en vue de prendre en charge les enjeux environnementaux et sociaux. Ce type de comportement n'est pas spontané car il est coûteux, difficile à garantir et souvent peu récompensé sur le marché – malgré les incitations émanant des acteurs publics et des changements de comportements des consommateurs (Chanteau, 2011).

L'engagement des entreprises dans la responsabilité sociale des entreprises s'explique d'une part, par la recherche d'une meilleure image auprès des consommateurs, des acteurs publics et de la société civile, et d'autre part par l'amélioration de la compétitivité d'une entreprise. De fait, cet engagement implique de nouer des collaborations avec des parties prenantes, et en particulier les organisations de la société civile. En effet, les entreprises qui démontrent un engagement plus fort envers la RSE auront une plus grande propension à collaborer avec les ONG que celles qui ont un engagement plus faible (Den Hond et al., 2012). Les interactions avec les parties prenantes externes peuvent fournir à une entreprise plusieurs types de ressources (Pfeffer et Salancik, 1978 ; Barney, 1991): une expertise et un savoir-faire, de la légitimité et de la réputation, des formes nouvelles de communication, et un accès préférentiel aux engagements ou au soutien d'acteurs importants en dehors de l'entreprise. Le choix de s'engager dans la RSE est lié aux vulnérabilités et aux opportunités spécifiques de l'entreprise (Smith, 2003). Les pressions extérieures sont aussi une incitation importante à la collaboration : l'aggravation des crises environnementales (Gray, 1989) ou la situation politique et économique (Dütting et Sogge, 2010). Le contexte de la RSE met ainsi davantage en exergue le fait que les ressources et compétences ne sont pas nécessairement détenues par une entreprise, mais résident aussi dans ses relations collaboratives avec d'autres organisations (Dyer et Singh, 1998).

Les travaux sur la responsabilité sociale des entreprises soulignent le rôle des grandes institutions internationales (Banque mondiale, Union européenne, OIT, etc.), des ONG locales, des gouvernements des pays en développement (Elyachar, 2003, 2005), ainsi que des milieux d'affaires (Maurer, 2012), dans des projets de développement. L'étude des systèmes sectoriels de production et d'innovation montre la pertinence de la prise en compte des interactions marchandes et non marchandes pour comprendre les mutations du système industriel (Bianchi et Labory S, 2013, Laurent et Du Tertre, 2008).

Dans plusieurs cas, la collaboration inter-organisationnelle s'effectue entre les entreprises et des organisations sans but lucratif, comme les associations. Les organisations associatives sont des entreprises privées collectives, sans but lucratif (du fait de l'absence ou la limitation de la rémunération individuelle du capital social) mais qui participent néanmoins aux échanges marchands et sont pleinement ancrées dans notre système productif. Dans plusieurs champs, elles participent à l'émergence et la structuration de nouvelles filières (textile, énergie, déchets, éco-construction) avec la spécificité d'intégrer des innovations technologiques comme des innovations sociales (Klein et al., 2014). Dès lors, elles participent à la prise en compte des dimensions économiques, environnementales et sociales auxquelles doit répondre la transition énergétique.

L'économie industrielle et les Sciences de gestion ont depuis plusieurs années mis en lumière les relations de coopération entre les firmes, entre autres dans des clusters, districts industriels

ou pôles de compétitivité (Forest et Hamdouch, 2009 ; Davis, 2016). Alors que les relations inter-firmes sont le plus souvent analysées sous l'angle de la théorie des contrats et de ses développements récents, il s'agit dans cette recherche d'interroger les facteurs de réussite de la collaboration inter-organisationnelle. Le cas particulier des filiales communes internationales a donné naissance à une littérature relativement abondante. Le succès de ces filiales a été étudié à travers des dimensions comme la survie des filiales, la performance financière, la satisfaction des partenaires, l'atteinte des objectifs et l'apprentissage (Ren et al., 2009). Les facteurs de performance analysés couvrent des thèmes vastes comme l'engagement, le contrôle et le pouvoir, la confiance, la justice, le conflit et sa résolution, la coopération, la distance culturelle et des objectifs partagés (Ren et al., 2009). Le degré de diversité entre les organisations partenaires influence la pérennité de ces collaborations : une variété « moyenne » et un équilibre entre les partenaires impactent positivement la survie de filiales communes multi-partenaires (Mohr, Wang et Goerzen, 2016).

1.2. Du choix d'une méthode d'analyse et des opportunités de la QCA

Les défenseurs du « paradigme » quantitatif ont toujours mis en avant le critère de scientificité de leur démarche (Bergadaà et Nyeck, 1992). Certes, à travers l'énumération des objets et de la fréquence d'apparition des phénomènes (Grawitz, 2001), seules les méthodes quantitatives permettent de conclure sur la possibilité de généralisation statistique des résultats. Mais la collecte de données quantitatives sur un grand nombre d'observations peut être difficilement réalisable en Sciences de gestion, surtout lorsque l'unité d'analyse observée se situe au niveau organisationnel ou inter-organisationnel. Au-delà, les méthodes qualitatives possèdent des qualités intrinsèques riches qui ont contribué à leur développement au fil des décennies. Elles décrivent les qualités essentielles des objets étudiés plutôt que de compter des fréquences. Souvent utilisées au sein d'études de cas, elles permettent de décrire les phénomènes de manière riche et contextualisée (Yin, 1989). Accusées par les chercheurs « quantitatifs » de ne pas être scientifiques (Bergadaà et Nyeck, 1992), les méthodes qualitatives se sont structurées et les démarches d'analyse ont été précisées (Eisenhardt, 1989, Gioia, Corley et Hamilton, 2012). Mais ces démarches concernent la rigueur de l'analyse des données et la validité interne des résultats, la généralisation statistique des résultats reste impossible. Par contre, cette méthode constitue une nouvelle approche pour étudier la causalité entre des variables (Wagemann et al., 2016). La quantification des données qualitatives apparaît comme une perspective intéressante pour faire face à ces différentes difficultés, et tenter de combiner les vertus de ces deux catégories de méthodes. Nous argumentons que le choix (ou non) d'une méthode de quantification de données qualitatives doit se faire en tenant compte de trois paramètres (cf tableau 1).

Méthodologie (et exemples de logiciels)	Taille de l'échantillon	Données	Codage	Modes de quantification des données
(Crisp-set) QCA (fs/QCA)	Petite à élevée, mais principalement intermédiaire	Variables binaires	Encodage de données qualitatives en variables binaires	En amont du traitement des données pour permettre l'analyse avec l'algèbre booléenne
Analyse de discours (Sphinx, Alceste)	D'un cas unique à un nombre élevé de cas	Mots / lexèmes	Pas de pré codage par le chercheur	L'analyse vise à compter les occurrences des lexèmes, les regroupe, les place sur des axes
Analyse de contenu (NVivo, Atlas.ti)	A partir d'un cas unique	Sens / contenu	Codage qualitatif	En aval du traitement, on peut compter les fréquences d'occurrence des catégories, le nombre de cas concernés par thème...

Tableau 1: Méthodes de quantification des données qualitatives en analyse de cas

La démarche de Ragin a été de consolider la rigueur de l'analyse comparative des données qualitatives en intégrant des lois statistiques (minimisation de Booléenne, traitement des cas logiques) (Chanson et al, 2005). La méthode QCA permet alors de traiter des échantillons intermédiaires¹ tout en gardant la rigueur scientifique et en interrogeant toujours les variables choisies, leur combinaison et les écarts par rapport aux résultats attendus (Chanson et al, 2005). Elle permet de comprendre la diversité des phénomènes grâce à l'hétérogénéité causale, à la différence de la linéarité monolithique et causale souvent induite par les méthodes quantitatives (Chanson et al., 2005). Elle permet ainsi de conclure sur les conditions nécessaires ou suffisantes à tel ou tel résultat.

La méthode QCA a été fondée par Ragin (1987) dans l'objectif de comparer de manière rigoureuse des « cas » dans les Sciences politiques comparées. D'abord restée confidentielle, le nombre de recherches utilisant cette méthode a fortement cru à partir de la période 2003-2005, lorsque la méthode a progressivement été adoptée dans d'autres champs disciplinaires comme la sociologie et les Sciences de gestion (Rihoux et al., 2013). Parallèlement, comme les autres méthodes de quantification des données qualitatives, la méthode QCA a bénéficié du développement de logiciels informatiques ; le plus connu et ancien est fs/QCA, co-développé par Ragin lui-même. Aujourd'hui complété et concurrencé par plusieurs autres logiciels, il correspond à une des variantes de la méthode. La QCA a donné lieu à de profondes évolutions au fil des années. L'approche initiale, appelée « crisp-set QCA » et basée sur des variables binaires a été complétée depuis par les variantes « fuzzy-set QCA » et « multi-value QCA » dans lesquelles les variables peuvent prendre plus de deux valeurs (Rihoux et al., 2014 ; Haesebrouck, 2016). Ces dernières sont difficilement envisageables sans le recours à un logiciel d'analyse, qui n'est pas indispensable pour l'approche initiale.

¹ Selon Rihoux et al. (2014) les échantillons intermédiaires comportent entre dix et cinquante cas.

Plusieurs facteurs justifient le choix d'utiliser la méthode QCA par rapport à d'autres méthodes alternatives. Ces éléments sont indispensables pour valider la cohérence de cette méthode par rapport au matériau disponible.

Premièrement, la méthode QCA est particulièrement appropriée dans des travaux qui interrogent les relations causales complexes et des interactions multiples (Fiss, 2011). La valeur ajoutée de cette méthode est de comprendre la façon dont les causes se combinent pour créer un résultat et propose alors de modéliser des configurations organisationnelles (Fiss, 2011). Par-delà sa capacité à systématiser la comparaison d'un nombre de cas « intermédiaire », la méthode QCA fournit une alternative aux méthodes quantitatives classiques par la comparaison de cas dans des configurations multiples. Elle étudie la diversité et l'hétérogénéité des cas en fonction de différentes conditions et de contextes causalement pertinents.

Deuxièmement, le choix peut être guidé par l'unité d'analyse. Dans les Sciences de gestion, l'approche par l'analyse des cas est historique. Mais, qu'est-ce qu'on entend par un « cas » ? En sociologie, par exemple, des recherches prenant comme unité d'analyse l'individu collectent des données sur de (très) grands échantillons qui peuvent ensuite être étudiés avec des méthodes quantitatives « classiques ». Des approches beaucoup plus « macro » prennent comme unité d'analyse un pays, par exemple en science politique ou en économie. En Sciences de gestion, les unités d'analyse vont de l'individu (notamment en marketing et en gestion des ressources humaines) au pays (par exemple en management cross-culturel comparé), en passant par des équipes, des projets et, notamment, des organisations. C'est dans ce contexte que la méthode QCA a commencé à être transposée et utilisée en Sciences de gestion (Curchod, 2003 ; Chanson et al., 2005).

Troisièmement, les possibilités d'accès aux données, les méthodes de collecte utilisées et le nombre et la variété de sources (entretiens, observations, documents...) étudiés, la richesse des données recueillies au sein d'un cas varie fortement. La quantification des données qualitatives entraîne une forte simplification, et réduit la richesse et la variété des représentations recueillies – notamment dans la variante initiale de la QCA, (« crisp-set QCA »), utilisant des variables binaires et pouvant être réalisée « manuellement ». En conséquence, plus les données sont riches, plus la quantification simplifie ; en revanche, pour des données plus basiques, la quantification présente davantage d'avantages que de limites. Les chercheurs rencontrent souvent des difficultés d'accès aux données pour étudier les organisations, qui est pourtant le niveau d'analyse privilégié dans des champs comme la stratégie (Chanson et al., 2005). Les méthodes de quantification des données qualitatives peuvent également contribuer à faire face à la difficulté que rencontrent les chercheurs à restituer des données qualitatives et leurs processus d'analyse dans les espaces contraints définis par les revues académiques. Elle permet donc un dialogue entre les utilisateurs des méthodes quantitatives et celles qualitatives.

Quatrièmement, la taille de l'échantillon et de la population totale influence le choix de la méthode. La méthode QCA a initialement été développée pour des échantillons de taille petite à intermédiaire, idéalement exhaustifs de la population totale (Chanson et al., 2005). A partir d'une trentaine d'observations, plusieurs méthodes d'analyse quantitative s'offrent aux

chercheurs, comme les régressions ou les modèles d'équations structurelles avec PLS. A l'inverse, pour un faible nombre de cas, la quantification des données qualitatives risque de comporter plus d'inconvénients que d'avantages, car la simplification des données n'est pas compensée par la possibilité d'énumérer des fréquences d'apparition des phénomènes. C'est précisément entre ces deux ordres de grandeurs que la méthode QCA est particulièrement appropriée, pour un nombre intermédiaire de cas (Rihoux et al., 2014), se situant approximativement entre une dizaine et une trentaine voire une cinquantaine. La méthode QCA permet donc bien d'analyser un nombre de « cas » intermédiaire, au-delà du cas unique ou d'un petit nombre de cas, mais en deçà d'un nombre de cas pouvant être analysé à l'aide de méthodes quantitatives « classiques », ce qui soulève des difficultés particulières. Aujourd'hui, la méthode QCA est également utilisée pour des grands échantillons (Rihoux et al., 2013), mais sa pertinence par rapport à des méthodes concurrentes est alors davantage sujette à débat.

Cinquièmement, la nature et les possibilités de codage des variables guident le choix méthodologique. La manière de quantifier les données qualitatives est fonction du schème de codage retenu. L'unité d'analyse retenue pour le codage des données qualitatives peut aller du mot (analyse de discours) à l'interprétation du sens exprimé dans un ensemble de phrases (analyse de contenu) (Allard-Poesi, 2003). Les logiciels d'aide à l'analyse de données qualitatives sont généralement spécialisés dans l'une ou l'autre de ces logiques, avec des logiciels comme Sphinx ou ALCESTE pour l'analyse des discours ou des représentations, et NVivo ou Atlas.ti pour l'analyse de contenu, pour ne citer que les plus connus. La méthode QCA compare des « cas » et correspond donc davantage aux recherches s'inscrivant dans des logiques d'analyse de contenu.

La méthode QCA s'appuie sur un processus itératif entre les cas et les variables. L'analyse des données implique des retours sur les données qualitatives et les cas étudiés (Legewie, 2013). Pendant ce processus, il est fréquent d'ajouter ou de supprimer des cas en fonction des combinaisons de conditions ou de conséquences obtenues (Wagemann, 2007).

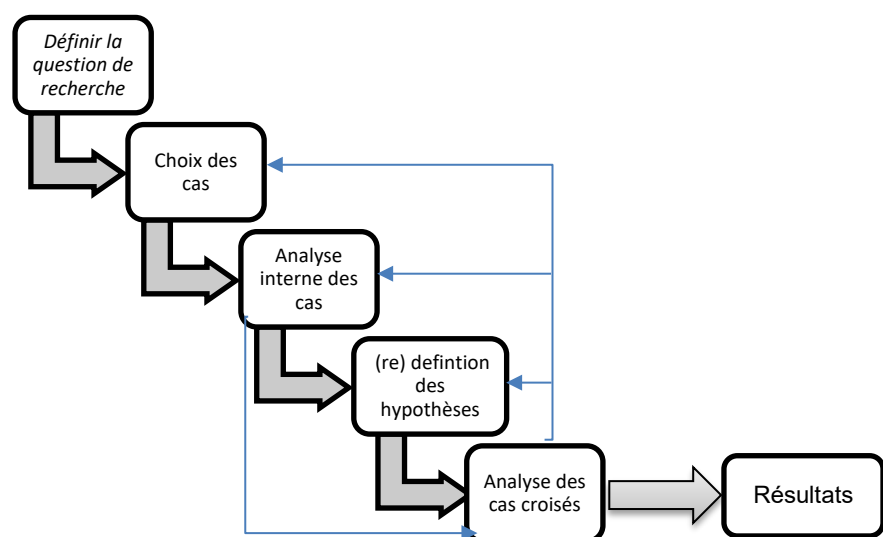


Figure 1: Le processus d'analyse de données selon la méthode QCA (Source : traduit de Legewie, 2013)

Comme d'autres méthodes de quantification des données qualitatives (par exemple l'analyse textuelle statistique avec la méthode ALCESTE ; Artis et Cornée, 2016), cette démarche nécessite une très bonne connaissance des cas. En effet, à la différence des méthodes quantitatives, ces méthodes imposent des allers-retours réguliers entre l'analyse et les données. Dans le cas d'ALCESTE, le logiciel ne prend pas en compte le sens, propre ou figuré, des formes textuelles utilisées. Ainsi, ALCESTE propose des pistes de réflexion et d'interprétation pour le chercheur pour distinguer et interpréter les classes, et lui laisse la possibilité de ne pas traiter les phénomènes anormaux, difficilement interprétables ou ne correspondant pas à ses hypothèses (Delavigne, 2004). Dans la méthode QCA, le traitement des cas contradictoires et l'étude des combinaisons dites logiques sont autant d'étapes d'interpellation du chercheur et de vérification de la cohérence du cadre analytique.

A l'inverse de méthodes où le chercheur ne maîtrise pas le processus, semblable à une boîte noire, la méthode QCA oblige le chercheur à maîtriser l'ensemble du processus d'analyse. La codification binaire des variables demande au chercheur un travail minutieux de déconstruction des cas et d'arbitrage à justifier théoriquement.

Les études de cas qualitatives sont souvent associées à des designs de recherche exploratoires (Bergadaà & Nyeck, 1992 ; Yin, 1989) et les études quantitatives sur des grands échantillons visent souvent le test d'un modèle théorique préexistant. La méthode QCA peut être utilisée aussi bien dans une logique inductive que déductive. Dans une visée déductive, les différentes conditions supposées mener à un résultat sont définies, et leur impact est observé dans les résultats. Dans une visée inductive et exploratoire, s'il est néanmoins nécessaire de définir le résultat analysé, le chercheur peut poser autant de conditions possibles que les données le permettent, afin de faire émerger parmi ces variables les conditions nécessaires ou suffisantes au résultat. La méthode QCA implique le regroupement des cas dans des catégories homogènes, définies par des caractéristiques précises. Le nombre de catégories possibles est de 2^x : un critère définit deux catégories, deux critères quatre catégories, trois critères huit catégories, et ainsi de suite. Quant au nombre de conditions à retenir, il existe principalement deux possibilités : soit, se limiter à un nombre de catégories nettement inférieur au nombre de cas observés, afin d'aboutir à quelques sous-ensembles de cas, soit, inclure plus de variables mais accepter alors que les cas observés se « répartissent » en de nombreuses catégories sans former des sous-ensembles. Un design de recherche déductif paraît davantage cohérent avec un petit nombre de conditions, alors que dans une visée exploratoire, il est pertinent d'inclure un grand nombre de conditions. Mais il est également possible d'explorer les données en comparant plusieurs « modèles » avec un faible nombre de conditions.

2. APPLICATION DE LA MÉTHODE QCA À LA COLLABORATION INTER-ORGANISATIONNELLE

Notre recherche prend comme unité d'analyse les collaborations entre organisations de natures diverses. En s'appuyant sur un système de saturation², nous avons trouvé trente collaborations, un nombre de cas pour lequel l'analyse QCA est très pertinente. Notre recherche se situe bien dans la logique de l'analyse de contenu, appropriée pour QCA.

Grâce à la méthode QCA, nous avons pu explorer trente cas constitués de données qualitatives et identifier les facteurs de succès de la collaboration inter-organisationnelle entre les grandes entreprises et des organisations sans but lucratif.

2.1. Sélection des cas et collecte des données qualitatives

Nous avons construit la base de données des cas en identifiant des projets d'énergies renouvelables à travers le monde. Les informations que nous avons collectées sont des données secondaires sur les cas, diffusées sur internet. Nous avons inclus des technologies énergétiques diverses : l'éolien, le solaire, l'hydraulique et la géothermique.

Les informations ont été collectées à partir des rapports écrits par les instances internationales investies dans le champ (par exemple : banques publiques d'investissements, agences internationales des énergies renouvelables, agences gouvernementales). Nous avons trouvé 30 cas de collaborations inter-organisationnelles rentrant dans nos critères. De façon systématique, nous avons identifié :

- les parties prenantes engagées dans le projet : entreprises, associations et ONG, centres de recherche, agences gouvernementales, financeurs ;
- les caractéristiques économiques du projet : localisation, date de démarrage du projet, durée de l'implémentation du projet, nombre de travailleurs, coût du projet ;
- les caractéristiques technologiques du projet : technologie, quantité d'énergie produite ;
- les caractéristiques de la situation énergétique du pays : quantité d'énergie consommée, tension sur le marché de l'énergie, dépendance vis-à-vis d'une source d'énergie fossile spécifique (pétrole, gaz, charbon...) ;
- l'acceptation du projet : avis des populations locales sur le projet (selon les médias).

2.2. Analyse QCA des données

Nous avons donc suivi les différentes étapes de la méthodologie QCA (Rihoux et Ragin, 2009) : i) sélection et codage des informations sur les cas sélectionnés, ii) construction du tableau de vérité en utilisant la dichotomisation des données, iii) vérification de la robustesse

² Nous avons complété notre base de cas en ajoutant tous les projets nouveaux que nous trouvons et qui respectaient nos critères (diversité organisationnelle, et énergies renouvelables).

des données (résolution des cas contradictoires, processus de minimisation booléenne et prise en compte des combinaisons dites « logiques ») et iv) interprétation des résultats.

Nous avons fait le choix de nous limiter à trois conditions équivalant à huit catégories. Ajouter une condition supplémentaire aurait créé 16 configurations théoriques, avec en moyenne de moins de deux cas par catégorie. Nous avons donc opéré un choix difficile car simplificateur pour retenir les trois conditions nous semblant les plus importantes au vu des données disponibles :

1. La présence d'incitations aux énergies renouvelables et à l'innovation : incitations publiques (réductions de taxes, financement dédié, etc.) et incitations émanant de la société civile. Par exemple, Singapour a investi plus de 570 millions de dollars américains depuis 2011 dans les incitations, en particulier dans les solutions d'énergie solaire (ASEAN Briefing, 2015). Les incitations émanant des politiques publiques telles que les réglementations, les subventions, permis négociables, ou la réduction des taxes ont un effet (direct ou indirect) positif sur l'innovation environnementale (Depret & Hamdouch, 2009).
2. La présence d'un conflit sur les ressources naturelles : l'exploitation des énergies renouvelables peuvent générer des conflits dans la gestion et l'usage des ressources naturelles. En effet, Ostrom (1990) met en lumière que les ressources renouvelables sont davantage sujettes à une rareté qu'à une abondance, et leur exploitation peut mener à des conflits entre les utilisateurs. Plusieurs études sur la gestion de l'eau cristallisent ces conflits dans les choix possibles d'usages de la ressource. Ces conflits sont aussi présents dans la gestion du foncier entre des usages alternatifs (usage agricole, production d'énergie ou immobilier).
3. La dimension internationale de la collaboration inter-organisationnelle. Par-delà ces éléments, la dimension internationale de la collaboration est un facteur majeur contribuant aux différences entre les acteurs (Mohr et al., 2016). Que ce soit au niveau culturel, linguistique, institutionnel, politique..., la nature internationale de la collaboration rajoute aux différences existant de toute façon entre les partenaires dans les cas que nous avons retenus, car il s'agit d'un côté d'entreprises et de l'autre côté d'organisations à but non lucratif.

Nous avons déterminé si les projets sont réussis à partir de trois critères liés à la contribution du projet à la transition énergétique :

- La quantité d'énergie produite par rapport à la capacité de production du pays,
- La durée de mise en œuvre du projet, en la comparant au temps moyen par mégawatt nécessaire dans des technologies équivalentes,
- La contribution du projet à un approvisionnement énergétique durable du pays d'implantation. Cet indicateur est composé de trois variables : la contribution du projet à la diversification des sources de production énergétique, l'impact en termes d'environnement et de réduction des émissions de CO², et la capacité productive de la technologie en réponse aux besoins locaux.

Nous avons déterminé le succès du projet à partir de la combinaison de ces indicateurs (succès = 1, échec = 0) : seuls les projets obtenant des « 1 » pour au moins deux des trois

critères ont été considérés comme des succès. Cette détermination multicritère du succès a pour conséquence de classer des projets en situation d'échec malgré leur production d'énergies renouvelables. Les projets ne sont en « échec » que de façon relative aux autres : ils ont nécessité plus de temps, la production d'énergie est faible ou l'impact environnemental moins marqué.

Nous avons renseigné ces facteurs à partir des données collectées au préalable, et nous les avons codés de façon dichotomique (tableau en annexe) pour ensuite construire la table de vérité.

Configuration	Incitations	Conflits	International	Cas	Nombre de cas	Résultat (succès ou non)
1	0	0	0	4	1	Positif
2	1	0	0	3,11,12,13,27	5	Positif avec 2 cas contradictoires
3	0	1	0	-	0	-
4	1	1	0	25	1	Négatif
5	0	0	1	6,8,9,21,15	5	Positif avec 1 cas contradictoire
6	1	0	1	1,5,7,10,14,18,19,20,24,26,28,29	12	Positif avec 2 cas contradictoires
7	0	1	1	2,23	2	Négatif
8	1	1	1	16,17,22,30	4	Négatif avec 1 cas contradictoire

Tableau 2 : Table de vérité des facteurs de succès des 30 cas d'études

2.3. Résultats de l'analyse CQA

Le tableau 2 présente les huit configurations possibles à partir de nos trois variables ; sept configurations peuvent être observées dans les données. La configuration numéro 6 est la plus fréquente. Elle se caractérise par des projets internationaux, n'ayant pas généré de conflit et localisés dans des pays où il existe des incitations pour les énergies renouvelables (12 projets). Nous avons cependant 2 cas contradictoires dans cette configuration (cas 26 et 29). Les configurations 2 et 5 présentent cinq cas chacune, la configuration 5 comporte un cas contradictoire, et la configuration 2 en possède deux. La configuration 5 regroupe les cas qui sont internationaux, n'ayant pas généré de conflit et sans incitation ; et la configuration 2 rassemble les cas locaux, sans conflit et avec des incitations. La configuration 8 est intéressante car elle regroupe les projets qui ont échoué, à l'exception d'un cas contradictoire. Paradoxalement, la présence simultanée de trois facteurs aboutit à des projets qui ont échoué. Les cas contradictoires ont été réévalués à partir des données disponibles permettant alors de résoudre ces contradictions. Par exemple, le cas 8 a été classé comme infructueux car la durée de mise en œuvre du projet est supérieure à la moyenne (1,7 MW par mois contre un temps moyen pour les projets solaires de 4 MW / mois) et que la production est considérée comme faible (production de 20,5MW par rapport à une production totale d'électricité en Chine de 1223GW). Ces deux critères ont conduit à classer ce cas comme infructueux. Cependant, ce projet s'appuie sur une technologie solaire alors que la production en Chine est

principalement fossile et hydroélectrique. Il participe donc à la diversification des sources d'énergie dans un contexte d'augmentation des besoins et de pression internationale (signature des accords COP21). Au vue de ces compléments, nous avons finalement considéré ce cas comme un succès. La même procédure de résolution des contradictions a été réalisée dans toutes les configurations.

Quatre configurations sont possibles pour arriver au succès. Dans toutes ces configurations, il y a absence de conflits sur les ressources naturelles. Ce critère est donc une condition suffisante pour un résultat positif.

La configuration la plus observée montre que deux relations causales sont possibles :

- S'il n'y a pas de conflit sur les ressources, alors le projet sera un succès ;
- Si le projet est international et qu'il existe des incitations, alors il sera un succès.

En reprenant les normes de la notation de l'algèbre booléenne (Chanson et al., 2005, Ragin, 1987), la formulation est la suivante :

$$\text{SUCCES} = \text{INCITATIONS} \cdot \text{INTERNATIONAL} \cdot \text{conflit}$$

Cette configuration représente 12 cas de notre population totale, et 12 cas de succès sur les 20 cas de succès (soit 60%).

Cependant, la répartition homogène des cas entre les configurations 2 et 5 laisse une indétermination entre les facteurs « incitations » et « international ». Sur les 23 cas où des mesures incitatives étaient disponibles, neuf cas n'ont pas abouti. Sur les 23 projets avec des acteurs de différents pays, huit n'ont pas réussi. Dès lors, il semble que les incitations sont une condition suffisante. Cependant, la différenciation entre ces deux conditions (international et incitation) est à approfondir. La méthode QCA permet de mettre en lumière la diversité des relations causales possibles entre les conditions et le résultat : il existe alors plusieurs modalités pour aboutir à une situation de succès.

La prise en considération des cas logiques permet de valider la robustesse de la méthode et du cadre conceptuel choisi pour cette étude (Carsten & Schneider, 2012). Sur les huit configurations possibles, une configuration (3) n'a pas été observée, elle correspond au cas logique suivant : une configuration d'échec en présence de conflit sur les ressources, sans dimension internationale ni incitation aux énergies renouvelables. Nous avons conclu que la condition « absence conflits » conduit à un résultat positif.

Une autre perspective pour le travail futur serait de comparer la collaboration multi-acteurs avec la collaboration interentreprises afin de remettre en question les différences entre ces différents types de collaboration.

3. DISCUSSION MÉTHODOLOGIQUE

Sur la base de l'exemple de notre recherche, nous revenons maintenant sur des aspects clés de la méthode QCA et leur enjeu pour les chercheurs utilisant cette méthode : le nombre de conditions retenues, la nature binaire des variables, l'analyse de causalités possibles plutôt que de probabilités, et la généralisation des résultats obtenus.

3.1. Le nombre de variables (« conditions ») intégrées

Une des limites de la méthode QCA est l'augmentation exponentielle du nombre de configurations possibles en fonction du nombre de critères retenus. L'ajout d'une variable doublant le nombre de configurations, trois critères déterminent huit configurations possibles, 4 critères, 16, etc. Principalement deux logiques d'utilisation de la QCA coexistent quant au choix du nombre de conditions : soit, objectif est d'aboutir à des catégories incluant plusieurs cas et, ainsi, à une ébauche de typologie (mais il faut alors limiter le nombre de conditions); soit, le chercheur souhaite explorer la pertinence d'un maximum de variables quitte à ce que chaque catégorie ne soit illustrée que par un ou quelques cas, ou aucun. Si pour un faible nombre de variables, une analyse manuelle est assez simple à réaliser, la multiplication des variables peut justifier le recours à des logiciels. 61% de 313 recherches publiées entre 1984 et 2011 et utilisant la méthode QCA mobilisent plus de quatre variables (Rihoux et al., 2013) – mais ces études datent en grande majorité d'après 2004, période à laquelle des logiciels informatiques d'analyse QCA se sont généralisés.

Nous avons retenu la première option : avec 30 cas collectés, nous nous sommes limitées à huit configurations, soit trois variables ou conditions. Nous pourrions enrichir nos résultats en rajoutant une variable supplémentaire. Ce développement nous permettrait de consolider l'importance du facteur « absence de conflit », d'affiner la compréhension des facteurs « international » et « incitations » et d'explorer une dimension supplémentaire, par exemple, la dimension « innovation » des technologies.

3.2. La nature binaire des variables

La codification binaire des variables est une des bases de la méthode dans sa version originelle (« crisp-set QCA »), réalisable sans recours à un logiciel spécifique. Dans notre recherche, nous avons exploré un type de collaboration inter-organisationnelle bien particulier, c'est-à-dire entre au moins une entreprise et au moins une organisation à but non lucratif, portant sur la mise en œuvre d'un projet dans le secteur des énergies renouvelables. Cette exploration se situant au niveau mondial, nous n'avions pas d'autre choix que de nous limiter à des données secondaires, collectées sur internet. Bien que nous ayons multiplié et croisé les sources et les documents, les données à notre disposition restent beaucoup plus pauvres que celles pouvant être recueillies in situ, par des entretiens semi-directifs et des observations directes, par exemple. Ceci est d'un côté regrettable – mais de l'autre côté, particulièrement compatible avec la méthode QCA car l'encodage des données en variables binaires appauvrit moins la richesse des données recueillies que dans d'autres méthodes.

L'encodage binaire des données représente une limite de la méthode, car beaucoup de dimensions étudiées correspondent à des variables continues. Dans notre étude, la mesure du succès des projets et sa quantification ont été un véritable défi. L'enjeu était de ne pas rester sur une mesure quantitative liée à la production énergétique, et de questionner la collaboration inter organisationnelle. Par exemple, l'utilisation du critère « durée de la mise en œuvre projet » avait aussi pour objectif de mesurer la qualité de la collaboration inter-organisationnelle. En effet, les formes d'organisations hybrides sont des modes de coordination entre agents alternatives au marché et à la concurrence qui soulèvent des critiques et des interrogations sur leur efficacité, notamment en termes de temps (Ménard, 2003). Cet aspect de la collaboration inter-organisationnelle est donc une dimension essentielle mais cette variable est par définition une variable continue. C'est en réponse à cette limite qu'ont été développées les variantes « fuzzy-set » et « multi-value set » de la QCA, intégrant des variables non binaires. Mais ce débat va au-delà de l'opposition de variables binaires et variables continues. Adopter la méthode QCA équivaut à entrer dans une logique de typologies (Fiss, 2011), à questionner si un cas fait ou non partie d'un type, et non pas de mesurer des intensités. Les typologies forment un pilier central du management stratégique et de la théorie des organisations (Fiss, 2011), et certaines sont extrêmement populaires au sein de la discipline. La binarité de variables bien choisies permet potentiellement d'en développer des nouvelles.

3.3. Une diversité causale plutôt que la probabilité d'un modèle unique

Avant même d'entamer une analyse basée sur l'algèbre booléenne, l'étape intermédiaire de la table de vérité comporte selon nous des vertus pour les recherches sur études de cas. Elle oblige les chercheurs à identifier clairement les principaux éléments de leur modèle théorique (émergeant ou pré-existant), de différencier des causes possibles des conséquences dans la formulation de leur problématique, et d'interroger systématiquement tous les cas quant à la présence ou absence de chacune de ces variables. La présentation sous forme d'un tableau (comme la table de vérité de la méthode QCA) a l'avantage de la simplicité et de la clarté. Certes, un tel tableau de synthèse des résultats peut aussi être utilisé à des seules fins de présentation, sans nécessairement chercher des causalités et sans donner lieu à une analyse quantitative subséquente (par exemple, Bartel-Radic, 2013).

La méthode QCA favorise ainsi la prise de distance du chercheur par rapport au terrain pour renforcer l'objectivité de l'analyse. De plus, dans certaines situations, cet éloignement par rapport à l'unité d'analyse peut être une solution alternative face à des difficultés physiques ou linguistiques d'accès aux données.

Présentée initialement comme une possibilité d'analyser les échantillons de taille petite à intermédiaire (Ragin, 1987), la méthode QCA est aujourd'hui souvent utilisée pour des échantillons de plus grande taille. 17% des 313 études analysées par Rihoux et al. (2013) ont des échantillons de plus de 100 observations, et d'autres méthodes d'analyse statistiques (régressions, équations structurelles, ...) auraient tout aussi bien pu être utilisées. Une majorité des études porte sur des échantillons de plus de 30 observations qui auraient pu être analysés avec la méthode PLS, par exemple. Plus qu'une méthode par défaut, la QCA est une manière différente d'aborder la causalité. Lorsque les modèles statistiques testent un modèle

causal unique, la QCA admet une diversité causale, c'est-à-dire la coexistence de plusieurs chemins parallèles permettant d'arriver à un résultat (Chanson et al., 2005).

Dans notre étude, si l'absence de conflits sur les ressources naturelles apparaît comme une condition suffisante, plusieurs combinaisons de configurations menant à des collaborations réussies coexistent.

Les cas contradictoires (nous en avons eu à traiter plusieurs dans notre étude), plutôt que des « observations atypiques » à mettre à l'écart, représentent une opportunité d'interroger systématiquement les cadres théoriques et les cas observés, permettant au chercheur d'objectiver sa recherche, en questionnant tout le temps ses postulats et ses hypothèses.

3.4. La généralisation des résultats obtenus

Le point précédent impacte directement la possibilité de généraliser les résultats obtenus. Un des enjeux de la méthode QCA réside dans les conditions de généralisation des résultats. L'objectif initial de la méthode est d'être « centrée sur les cas » (Ragin, 1987), ce qui est un avantage tout comme un inconvénient. La « dépendance des cas » a de manière récurrente été citée comme une limite de la QCA (Rihoux et al., 2014). Dans le cas où la recherche porte sur toute une population de cas, de manière exhaustive (un des critères initiaux de la QCA, cf. Chanson et al., 2005), la question de la généralisation ne se pose qu'en dehors de cette population.

Nous ne pouvons pas nous prononcer sur l'exhaustivité de notre échantillon, mais il est très probable que nous n'avons pas repéré l'ensemble des projets en énergies renouvelables existant sur la planète associant des entreprises et des organisations sans but lucratif – autrement dit, que l'échantillon n'est pas exhaustif. Ceci questionne la généralisation de nos résultats, du fait de l'absence de données.

Comme pour les recherches par études de cas en général, si une généralisation statistique est impossible, une généralisation analytique est envisageable et utile (Yin, 1989). Elle repose sur une description fine et détaillée du contexte des cas, qui doit donc nécessairement accompagner l'analyse QCA. En d'autres termes, si la « table de vérité » a été construite suite à la quantification de données qualitatives, le chercheur dispose des données riches qui permettent assez facilement de contextualiser et préciser les résultats obtenus grâce à l'analyse QCA. A l'inverse, si les variables analysées dans la QCA ont pu être récoltées autrement (dans des bases de données quantitatives pré-existantes, par exemple), ce point devrait s'avérer plus problématique. La combinaison de la méthode QCA avec d'autres méthodes d'analyse, qualitatives ou quantitatives, est pointée comme très souhaitable par certains chercheurs (Schneider et Wagemann, 2012). Elle est pourtant absente de 61,3% des recherches référencées par Rihoux et collègues (2013), dans lesquelles l'analyse QCA ne coexiste avec aucune autre méthode d'analyse des données.

CONCLUSION : PRÉCONISATIONS POUR L'UTILISATION DE LA MÉTHODE QCA EN SCIENCES DE GESTION

L'objectif de cette contribution fut de présenter et de discuter des apports et des limites de la méthode QCA pour des chercheurs travaillant avec des études de cas qualitatives. Nous avons choisi d'explicitier les étapes du processus QCA en questionnant systématiquement ses spécificités par rapport à d'autres méthodes de recherche. Nous avons discuté les choix qui s'imposent à tout chercheur qui souhaite utiliser cette méthode. Pour illustrer notre propos, nous nous sommes appuyés sur une recherche sur la collaboration inter-organisationnelle dans le champ des énergies renouvelables.

Face aux limites des méthodes qualitatives et quantitatives de collecte et de traitement des données, la méthode QCA offre des perspectives prometteuses. Elle permet d'établir des causalités entre des variables qualitatives pour des échantillons de cas de taille intermédiaire. Elle enrichit les approches qualitatives et quantitatives tout en assurant une rigueur statistique ; elle conserve la richesse des approches qualitatives tout en permettant une démarche de généralisation. Cette méthode est aussi un challenge pour le chercheur car elle l'oblige à systématiquement s'interroger sur ses cadres théoriques et ses présupposés. La maîtrise des données (de leur nature, collecte et traitement) et l'explicitation des différentes étapes du *research design* sont essentielles à la bonne utilisation de cette méthode. Pourtant, et malgré son fort développement récent, cette méthode reste confinée à quelques cercles de chercheurs initiés aux débats sur l'épistémologie des méthodes en Sciences sociales.

L'utilisation de QCA nous a permis de dégager un résultat important : la collaboration inter-organisationnelle dans les énergies renouvelables est un succès dans des contextes où il n'existe pas de conflit sur la gestion et l'usage des ressources naturelles. Les nouvelles approches sur la gestion des ressources naturelles développées par Elinor Ostrom et ses collaborateurs montrent que le conflit est omniprésent dans la gestion collective des ressources communes. Dans ce cadre, la fabrique des institutions est en elle-même « un processus difficile, chronophage, source de conflits » (Ostrom, 1990, p. 14). Elle théorise le conflit comme un problème cognitif dû aux différences d'interprétation, et présente donc la mise en place d'institutions au sein de la communauté comme un processus permettant l'interaction et l'interdépendance entre les acteurs, pouvant augmenter la confiance mutuelle et diminuer les conflits. Une des perspectives pour notre recherche serait alors d'étudier les institutions au sens d'Ostrom mises en place lors de ces collaborations réussies. Autrement dit, comment les collaborations avec les organisations à but non lucratif ont-elles permis d'éviter les conflits majeurs sur les ressources et faciliter ainsi l'implémentation plus rapide de projets plus significatifs dans les énergies renouvelables ?

La généralisation de ce résultat n'est pas évidente autant du fait de nos données que de la méthode utilisée. Le choix d'un nombre limité de variables et la codification binaire simplifient fortement les résultats. Mais la méthode QCA nous a aussi permis de discuter des différentes configurations et leur combinaison pour aboutir à une collaboration réussie sur un champ d'exploration très vaste. Même si nous n'avons pas pu réaliser directement d'observations in situ, la méthode QCA a permis d'enrichir le débat sur cette thématique.

ANNEXE : DESCRIPTION ET CODAGE DES PROJETS

	Cas	Techno- logie	Localisa-tion	Incita- tions	Con- flit	Inter- national	Suc- cess
1	Solar farm in ASYV	Solaire	Rouanda	1	0	1	1
2	Hydropower in Intibucá	Hydro	Honduras	0	1	1	0
3	Hydropower Nepal	Hydro	Népal	1	0	0	1
4	Solar Lantern Rental System Laos	Solaire	Laos	0	0	0	1
5	Salkhit wind farm	Eolien	Mongolie	1	0	1	1
6	Sapphire wind farm	Eolien	Pakistan	0	0	1	1
7	The Lesedi and Letsatsi projects	PV	Afrique du Sud	1	0	1	1
8	Hami project	CPV	Chine	1	0	1	0
9	Wind farm in Singida	Eolien	Tanzanie	0	0	1	1
10	Karadzhalovo solar park	Solaire	Bulgarie	1	0	1	1
11	Wind farm in Madhya Pradesh, Andhra Pradesh	Eolien	Inde	1	0	0	1
12	Energy efficiency	Efficience énergq.	Chine	1	0	0	0
13	IVANPAH	Solaire Th.	USA	1	0	0	1
14	Ouarzazate Solar Power Station "Phase one"	Solaire PV	Maroc	1	0	1	1
15	Bugoye Hydro Power	Hydro	Ouganda	0	0	1	1
16	The Lake Turkana Wind Power Project (LTWP)	Eolien	Kenya	1	1	1	1
17	Palo Viejo Hydro Plant	Hydro	Guatemala	1	1	1	0
18	San Jacinto-Tizate Geothermal Project	Géo- thermal	Nicaragua	1	0	1	1
19	Jasper Solar Photovoltaic Power Plant	Solaire PV	Afrique du Sud	1	0	1	1
20	Amanecer Solar CAP Power Plant	Solaire PV	Chili	1	0	1	1
21	Chaglla Hydroelectric Power Plant	Hydro	Perou	0	0	1	1
22	Ulubelu II Geothermal Power Plant	Géo- thermal	Indonesie	1	1	1	0
23	Dai Ninh Hydro Plant	Hydro	Vietnam	0	1	1	0
24	Burgos Wind Project	Wind	Philippines	1	0	1	1
25	Rampur Hydro Project	Hydro	Inde	1	1	0	0
26	Agua Prieta II	ISCC	Mexique	1	0	1	0
27	Augustin Fresnel 1, Targassonne	Solaire	France	1	0	0	0
28	Hassi R'mel	ISCC	Algerie	1	0	1	1
29	Khi Solar One	CSP	Afr. du Sud	1	0	1	0
30	Manchasol 1	Solaire Thermal	Espagne	1	1	1	0

BIBLIOGRAPHIE

- Allard-Poesi F. (2003), « Coder les données », in : Giordano Y. (dir.) *Conduire un projet de recherche-Une perspective qualitative*, EMS-Management et société, p. 245-290.
- Artis A. et Cornée S. (2016), « La face cachée de l'intermédiation financière : composition, traduction et mémorisation du savoir idiosyncratique dans la banque solidaire », *Systèmes d'Information et Management*, vol. 21, n° 3.
- ASEAN Briefing (2015), « Incentives For Renewable Energy Investment », Singapore, N.p.
- Barney J. (2001), « Resource-based theories of competitive advantage: A ten-year retrospective on the resource-based view », *Journal of Management*, vol. 27, n° 6, p. 643-650.
- Bartel-Radic A. (2013), « 'Estrangeirismo' and Flexibility: Intercultural Learning in Brazilian MNCs », *Management International*, vol. 17, n° 4, p. 239-253.
- Bergadaà M. et Nyeck, S. (1992), « Recherche en marketing: un état des controverses », *Recherche et Applications en Marketing*, vol. 7, n° 3, p. 23-44.
- Bianchi P. et Labory S. (2013), « Structural Transformations in Industry and Filières », *Revue d'économie industrielle*, n° 144, p. 177-199.
- Chanson G., Demil B., Lecocq X. et Sprimont P. A. (2005), « La place de l'analyse qualitative comparée en sciences de gestion », *Finance Contrôle Stratégie*, vol. 8, n° 3, p. 29-50.
- Chanteau J.-P. (2011), « L'économie de la RSE : éléments de méthode institutionnaliste », *Revue de la régulation*, n° 9.
- Curchod C. (2003), « La méthode comparative en sciences de gestion : vers une approche quasi-expérimentale de la réalité managériale », *Finance Contrôle Stratégie*, vol. 6, n° 2, p. 155-177.
- Davis J. P. (2016), « The Group Dynamics of Inter-Organizational Relationships: Collaborating with Multiple Partners in Innovation Ecosystems », *Administrative Science Quarterly*, vol. 61, n° 4, p. 621-661.
- Delavigne V. (2004), « Présentation d'Alceste », *Revue Texto!*, n°9, p. 324-329.
- DeMeur G. et Rihoux B. (2002), *L'Analyse Quali-Quantitative Comparée*, Academia Bruylant.
- Den Hond F., de Bakker F. G. et Doh J. (2015). « What prompts companies to collaboration with NGOs? Recent evidence from the Netherlands ». *Business & Society*, vol. 54, n° 2, p. 187-228.
- Depret M.-H. et Hamdouch A. (2009), « Quelles politiques de l'innovation et de l'environnement pour quelle dynamique d'innovation environnementale ? », *Innovations*, n° 29, p. 127-147.
- Dütting G. et Sogge D. (2010), « Building Safety Nets in the Global Politic: NGO collaboration for solidarity and sustainability », *Development*, vol. 53, n° 3, p. 350-355.
- Dyer J. et Singh H. (1998), « The Relational View: Cooperative Strategy and Sources of Interorganizational Competitive Advantage », *Academy of Management Review*, vol. 23, n° 4, p. 660-679.
- Eisenhardt K. M. (1989), « Building Theories from Case Study Research », *Academy of Management Review*, vol. 14, n° 4, p. 532-550.
- Elyachar J. (2003), « Mappings of power: the state, NGOs, and international organizations in the informal economy of Cairo », *Comparative Studies in Society and History*, vol. 45, n° 3, p. 571-605.
- Elyachar J. (2005), *Markets of dispossession: NGOs, economic development, and the state in Cairo*, Duke University Press Books.
- Fiss, P. C. (2011), « Building Better Causal Theories: A Fuzzy-Set Approach to Typologies in

Organization Research », *Academy of Management Journal*, vol. 54, n° 2, p. 393-420.

Forest J. et Hamdouch A. (2009), « Les clusters à l'ère de la mondialisation : fondements et perspectives de recherche », *Revue d'économie industrielle*, n° 128, p. 9-20.

Gioia D. A., Corley K. G. et Hamilton A. (2012), « Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology », *Organizational Research Methods*, vol. 16, n° 1, p. 15-31.

Grawitz M. (2001), *Méthodes des sciences sociales*, Dalloz.

Gray B. (1989), « Developing a Theory for Collaboration », in: Gray B. (dir.), *Collaborating: Finding common ground for multiparty problems*, New Jersey, Jossey-Bass, p. 226–245.

Greckhamer T., Misangyi V. F., Elms H. et Lacey R. (2008), « Using qualitative comparative analysis in strategic management research: An examination of combinations of industry, corporate, and business-unit effects », *Organizational Research Methods*, vol. 11, n° 4, p. 695-726.

Haesebrouck T. (2016) « The Added Value of Multi-Value Qualitative Comparative Analysis », *Forum: Qualitative Social Research*, vol. 17, n° 1, art. 12, p. 1-29.

Klein J. L., Laville J. L. et Moulaert, F. (2014), *L'innovation sociale*, Erès.

Laurent C. et Du Tertre C. (dir.) (2008), *Secteurs et territoires dans les régulations émergentes*, Paris, L'Harmattan.

Legewie N. (2013), « An introduction to applied data analysis with qualitative comparative analysis », *Forum: Qualitative Social Research*, vol. 14, n° 3, p. 1-45.

Maurer B. (2012), « Mobile money: Communication, consumption and change in the payments space », *Journal of Development Studies*, vol. 48, n° 5, p.589–604.

Ménard C. (2003), « Économie néo-institutionnelle et politique de la concurrence les cas des formes organisationnelles hybrides », *Économie rurale*, n°277-278, p. 45-60.

Mohr A. T., Wang C. et Goerzen A. (2016), « The impact of partner diversity within multiparty international joint ventures », *International Business Review*, vol. 25, n° 4, p. 883-894.

Ostrom E. (1990), *Governing the Commons: The evolution of institutions for collective action*, Cambridge University Press.

Ragin C.C. (1987), *The Comparative Method*, University of California Press.

Ragin C.C. (2000), *Fuzzy-Set Social Science*, University of Chicago Press.

Ren H., Gray B. et Kim, K. (2009), « Performance of International Joint Ventures: What Factors Really Make a Difference and How? », *Journal of Management*, vol. 35, n° 3, p. 805-832.

Reverdy T. (2014), *La construction politique du prix de l'énergie: sociologie d'une réforme libérale*, Presses de Sciences Po.

Rihoux B., Marx A. et Alamos-Concha P. (2014), « 25 années de QCA (Qualitative Comparative Analysis) : quel chemin parcouru ? », *Revue internationale de politique comparée*, vol. 21, n° 2, p. 61-79.

Rihoux B., Alamos-Concha P., Bol D., Marx A. et Rezsöhazi I. (2013), « From niche to mainstream method? A comprehensive mapping of QCA applications in journal articles from 1984 to 2011 », *Political Research Quarterly*, vol. 66, n° 1, p. 175-184.

Rihoux B. et Ragin C. (2009), *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*, New York, Sage.

Schneider C. Q. et Wagemann C. (2010), « Standards of good practice in qualitative comparative analysis (QCA) and fuzzy-sets », *Comparative Sociology*, vol. 9, n° 3, p. 397-418.

Schneider C. Q. et Wagemann C. (2012), *Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis*, Cambridge, Cambridge University Press.

- Seny Kan A. K., Adegbite E., El Omari S. et Abdellatif M. (2016), « On the use of qualitative comparative analysis in management », *Journal of Business Research*, vol. 69, n° 4, p. 1458-1463.
- Sims C. A. (2010), « But Economics Is Not an Experimental Science », *Journal of Economic Perspectives*, vol. 24, n° 2, p. 59-68.
- Smith N. C. (2003), « Corporate social responsibility: Whether or how? », *California management review*, vol. 45, n° 4, p. 52-76.
- Thomas J., O'Mara-Eves A. et Brunton G. (2014), « Using qualitative comparative analysis (QCA) in systematic reviews of complex interventions: a worked example », *Systematic reviews*, vol. 3, n° 1, p. 679-706.
- Wagemann C., Buche J., et Siewert M. B. (2016), « QCA and business research: Work in progress or a consolidated agenda? », *Journal of Business Research*, vol. 69, n° 7, p. 2531-2540.
- Yin R. (1989), *Case Study Research: Design and Methods*, Sage.

AUTEURS

Amélie Artis

Maître de Conférences en économie
Univ. Grenoble Alpes, Sciences Po Grenoble, CNRS, PACTE, F-38000 Grenoble, France

Anne Bartel-Radic

Professeure des universités en management international
Univ. Grenoble Alpes, Grenoble INP, CERAG, F-38000 Grenoble, France

Hezam Haidar

Doctorant en génie industriel
Univ. Grenoble Alpes, Grenoble INP, CNRS, G-SCOP, F-38000 Grenoble, France

Quantifying qualitative case study research with QCA: implications and process of the method from the example of a study on inter-organizational collaboration

ABSTRACT

This chapter discusses the quantification of qualitative case study research with the QCA method on the example of a study of inter-organizational collaboration. The study explores collaborations between companies and not-for-profit organizations implementing projects in the renewable energy sector, using secondary data. Qualitative comparative analysis permits to identify the success factors of these collaborations. The chapter details the opportunities and limits of this method of analysis, and its implications for researchers using qualitative case studies.

KEYWORDS: Inter-organizational collaboration, Not-for-profit organizations, QCA (Qualitative Comparable Analysis), Renewable energy sector

INTRODUCTION

For more than thirty years, there has been an opposition between quantitative and qualitative research methodologies in the field of management science. Beyond the fact that the two categories of methods answer different and complementary objectives (Bergadaà and Nyeck, 1992), the quantification of qualitative data can contribute in closing the "gap" that continues to separate them. One of the methods used to quantify qualitative data is QCA, Qualitative Comparative Analysis, also called comparative qualitative analysis (Ragin, 1987, Rihoux, Marx and Alamos-Concha, 2014). Although the increase in using QCA in management was predicted more than ten years ago (Chanson et al., 2005), and despite a significant growth since 2005 (Rihoux, Alamos-Concha, Bol, Marx and Rezsöhazy, 2013; Kan, Adegbite, El Omari and Abdellatif, 2016), researchers still largely ignore this method. Its interest lies in the analysis of the causal complexity that involves different combinations of conditions, capable of generating the same result on samples of intermediate size.

The objective of this chapter is not to re-describe the QCA method and its past or potential uses for the study of organizations, but to discuss the interest of this method among the different methods of quantifying qualitative data for a didactic purpose. This method, initiated in comparative political science, is increasingly used in sociology and management science, particularly in strategy (Wagemann et al., 2016, Rihoux et al., 2013) and in management (Seny Kan et al., 2016). The challenge for its founder, C. C. Ragin (1987), was to develop a social science method that articulates "case-centered" and "variable-centered" approaches (Rihoux

et al., 2014: 77). Even though other uses are possible, and sophisticated and computerized variants have been developed, we argue that the "basic version" of the QCA deserves to be better known by researchers collecting qualitative data in case studies.

We rely on a study that we conducted on inter-organizational collaborations, in which we used the QCA method for the analysis of collected data. Based on this study, we will discuss the opportunities and limitations of this method for social scientists. Through our example, the challenge is to present research protocol of a study to discuss the methodological choices faced by social scientists and the contributions of the QCA method compared to other methods. The originality is to explain the "unthinking" methodological and their "naive" techniques (Sims, 2010). In our example, we are interested in the success factors of collaborations between companies and non-profit organizations (associations, NGOs, research organizations and public) to implement a project in the renewable energy sector - for the vast majority. In some cases, it involves the construction of power plants using renewable energies (solar, hydro, wind and geothermal).

We will proceed as follow: in the first part, we will present the theoretical framework and the problematic of the study on the collaborations between organizations of different natures. In addition, we will question the opportunity to approach them with the QCA method compared to other methodologies. In the second part, we will present the empirical data collected on thirty cases of collaborations in detail, and their analysis with the QCA method. Finally, taking inspiration from this research example, we will discuss the contributions and limits of the QCA method for management science and propose recommendations for its further use.

1. CONTEXT OF THE RESEARCH: THEMATIC AND CHOICE OF ANALYSIS' METHOD

We will first briefly present the literature on the subject studied, the inter-organizational collaborations in the renewable energy sector, before addressing the interest of retaining the QCA among the various methodologies that can be considered in order to carry out a study on the topic.

1.1. Review of the literature on inter-organizational collaborations in renewable energy projects

Renewable energies are a sector of technological innovation implemented by large multinational companies. Often created by governments, large companies in the energy sector today have more and more injunctions to meet the challenges of the energy transition (Reverdy, 2014). Yet these companies are not significantly engaged in the transition. Few take financial risks in order to take care of environmental and social issues. This type of behavior is not spontaneous because it is expensive, difficult to guarantee and often not rewarded in the market - despite the incentives emanating from public players and changes in consumer behavior (Chanteau, 2011).

The commitment of companies to corporate social responsibility (CSR) can be explained on the one hand by the search for a better image among consumers, public actors and civil

society, and on the other hand by the improvement of the competitiveness of a company. In fact, this commitment involves collaborating with different stakeholders, especially civil society organizations. Indeed, companies that demonstrate a stronger commitment to CSR will have a greater propensity to collaborate with NGOs than those with lower engagement (Den Hond et al., 2012). Interactions with external stakeholders can provide a firm with several types of resources (Pfeffer and Salancik, 1978, Barney, 1991): expertise and know-how, legitimacy and reputation, new forms of communication, and preferential access to supports from key players outside the company. The choice to engage in CSR is linked to the specific vulnerabilities and opportunities of the company (Smith, 2003). External pressures are also an important incentive for collaboration: the worsening of environmental crises (Gray, 1989) or the political and economic situation (Dütting and Sogge, 2010). The context of CSR thus highlights the fact that resources and skills are not necessarily owned by a firm, but also reside in its collaborative relationships with other organizations (Dyer and Singh, 1998).

Work on corporate social responsibility highlights the role of large international institutions (World Bank, European Union, ILO, etc.), local NGOs, governments in developing countries (Elyachar, 2003, 2005), as well as business communities (Maurer, 2012), in development projects. The study of sectoral systems of production and innovation shows the relevance of taking into account market and non-market interactions to understand the changes in the industrial system (Bianchi and Labory S, 2013, Laurent and Du Tertre, 2008).

In many cases, inter-organizational collaboration occurs between businesses and non-profit organizations, such as associations. The associative organizations are collective private enterprises, not-for-profit (because of the absence or the limitation of the individual remuneration of the social capital) but which nevertheless participate in commercial exchanges and are fully anchored in our productive system. In several fields, they participate in the emergence and structuring of new sectors (textiles, energy, waste, eco-construction) with the specificity of integrating technological innovations as social innovations (Klein et al., 2014). Therefore, they participate in taking into account the economic, environmental and social dimensions to which the energy transition must respond.

The industrial economy and management sciences have for several years highlighted cooperations between firms, such as those in clusters, industrial districts or innovation centers (Forest and Hamdouch, 2009, Davis, 2016). While inter-firm relations are most often analyzed from the perspective of contract theory and its recent developments, this research seeks to question the success factors of inter-organizational collaboration. The special case of international joint subsidiaries has given rise to a relatively abundant literature. The success of these subsidiaries has been studied through different dimensions such as the survival of subsidiaries, financial performance, partner satisfaction, achievement of objectives and learning (Ren et al., 2009). The analyzed performance factors cover broad themes such as engagement, control and power, trust, justice, conflict and resolution, cooperation, cultural distance, and shared goals (Ren et al., 2009). The degree of diversity between the partner organizations influences the sustainability of these collaborations; a "medium" variety and a balance between the partners have a positive impact on the survival of joint multi-partner subsidiaries (Mohr, Wang and Goerzen, 2016).

1.2. The choice of analysis' method and the opportunities of QCA

Supporters of the quantitative model have always proposed a scientific criterion for their approach (Bergadaà and Nyeck, 1992). Admittedly, through the enumeration of things and the frequency of phenomena (Grawitz, 2001), only quantitative methods allow us to conclude on the possibility of statistical dissemination of results. However, the collection of quantitative data on a large number of observations can be difficult to achieve in management sciences, especially when the observed unit of analysis is at the organizational or inter-organizational level. Beyond that, qualitative methods have rich intrinsic qualities that have contributed to their development over the decades. They describe the essential qualities of the objects studied rather than counting frequencies. Often used in case studies, they make it possible to describe phenomena in a rich and contextualized way (Yin, 1989). Accused by "quantitative" researchers of not being scientific (Bergadaà and Nyeck, 1992), qualitative methods have been structured and analytical approaches have been specified (Eisenhardt, 1989, Gioia, Corley and Hamilton, 2012). But these approaches concern the rigor of the analysis of the data and the internal validity of the results, the statistical generalization of the results remains impossible. On the other hand, this method constitutes a new approach to study the causality between variables (Wagemann et al., 2016). The quantification of qualitative data appears as an interesting perspective to face these different difficulties, and to try to combine the virtues of these two categories of methods. We argue that the choice (or not) of a method for quantifying qualitative data must be made taking three parameters into account (see Table 1).

Methodology (software example)	Sample size	Data	Coding	Data Quantization Modes
(Crisp-set) QCA (fs/QCA)	Small to big, but mainly intermediate	Binary variables	Encoding qualitative data in binary variables	Upstream of data processing to enable analysis with Boolean algebra
Discourse Analysis (Sphinx, Alceste)	From a single case to many cases	Words / lexemes	No pre- coding by the researcher	The analysis aims at counting the occurrences of the lexemes, groups them together, places them on axes
Content Analysis (NVivo, Atlas.ti)	From a single case	Meaning / content	Qualitative coding	Downstream of the treatment, we can count the frequencies of occurrence of categories, the number of cases concerned by theme ...

Table 2: Quantification Methods for Qualitative Data in Case Analysis

Ragin's approach was to consolidate the rigor of the comparative analysis of qualitative data by integrating statistical laws (Boolean minimization, treatment of logical cases) (Chanson et

al, 2005). The QCA method then makes it possible to process intermediate samples³ while maintaining the scientific rigor and always interrogating the chosen variables, their combination and the deviations from the expected results (Chanson et al, 2005). It makes it possible to understand the diversity of phenomena thanks to the causal heterogeneity, unlike the monolithic and causal linearity often induced by quantitative methods (Chanson et al., 2005). It thus makes it possible to conclude on the necessary or sufficient conditions for this or that result.

The QCA method was founded by Ragin (1987) in order to rigorously compare "cases" in comparative political science. First kept confidential, the number of research cases using this method strongly increased from the period 2003-2005, when the method was gradually adopted in other disciplinary fields such as sociology and management sciences (Rihoux et al., 2013). At the same time, like the other methods of quantification of qualitative data, the QCA method benefited from the development of computer software; the oldest and most well-known is fs / QCA, co-developed by Ragin himself. Today completed by and in competition with several other software, it corresponds to one of the variants of the method. The QCA went through profound changes over the years. The initial approach, called "QCA crisp-set" and based on binary variables has since been supplemented by the "fuzzy-set QCA" and "multi-value QCA" variants in which variables can take more than two values (Rihoux et al., 2014, Haesebrouck, 2016). These are difficult to envisage without the use of analysis software, which is not essential for the initial approach.

Several factors justify the choice to use the QCA method over other alternative methods. These elements are essential to validate the consistency of this method with the available material. First, the QCA method is particularly appropriate in work that questions complex causal relationships and multiple interactions (Fiss, 2011). The added value of this method is to understand how causes combine to create an outcome and then propose to model organizational configurations (Fiss, 2011). Beyond its ability to systematize the comparison of "intermediate" sample of case studies, the QCA method provides an alternative to conventional quantitative methods by comparing cases in multiple configurations. It studies the diversity and heterogeneity of cases according to different conditions and causally relevant contexts. Second, the choice can be guided by the unit of analysis. In management science, the case analysis approach is historic. But what is meant by a "case"? In sociology, for example, when conducting research as a unit of analysis, the individual collects data on very large samples that can be studied in "classical" quantitative methods. Much more "macro" approaches take a country as a unit of analysis, for example in political science or economics. In Business Sciences, the units of analysis range from the individual (notably in marketing and human resources management) to the country (for example in comparative cross-cultural management), through teams, projects and, in particular, organizations. It is in this context that the QCA method has begun to be transposed and used in management science (Curchod 2003, Chanson et al., 2005).

³ According to Rihoux et al. (2014) the intermediate sample contains between ten and fifty cases.

Third, the possibilities of access to data, the collection methods used and the number and variety of sources (interviews, observations, documents ...) studied represents the richness of the data collected within a case, and that varies greatly. The quantification of qualitative data leads to a great simplification, and reduces the richness and variety of representations collected - especially in the initial variant of the QCA ("crisp-set QCA") using binary variables and can be performed "manually". As a result, the richer the data, the simpler the quantification; on the other hand, for more basic data, quantification has more advantages than limits. Researchers often encounter difficulties in accessing data to study organizations, which is the preferred level of analysis in fields such as strategy (Chanson et al., 2005). Qualitative data in quantification methods can also help to cope with the difficulty researchers encounter in restoring qualitative data and their analysis processes in constrained spaces defined by academic journals. Thus it allows a dialogue between users of quantitative and qualitative methods.

Fourth, the size of the sample and the total population influence the choice of the method. The QCA method was initially developed for small to intermediate sample sizes, ideally exhaustive of the total population (Chanson et al., 2005). From about thirty observations, several methods of quantitative analysis are offered to researchers, such as regressions or models of structural equations with PLS. Conversely, for a small number of cases, the quantification of qualitative data may have more disadvantages than advantages, since the simplification of the data is not offset by the possibility of listing frequencies of occurrence of the phenomena. It is precisely between these two orders of magnitude that the QCA method is particularly appropriate, for an intermediate number of cases (Rihoux et al., 2014), being approximately between ten and thirty or even fifty. The QCA method therefore makes it possible to analyze a number of intermediate "cases", beyond the single case or a small number of cases, but below a number of cases that can be analyzed using "conventional" methods, which raises particular difficulties. Today, the QCA method is also used for large samples (Rihoux et al., 2013), but then its relevance to competing methods is more open to debate.

Fifth, the nature and coding possibilities of the variables guide the methodological choice. The way to quantify the qualitative data depends on the chosen coding scheme. The unit of analysis used for the coding of qualitative data can range from the word (discourse analysis) to the interpretation of meaning expressed in a set of sentences (content analysis) (Allard-Poesi, 2003). Qualitative data analysis software is generally specialized in one or more of these logics, with software such as Sphinx or ALCESTE for discourse analysis or representations, and NVivo or Atlas.ti for content analysis. The QCA method compares "cases" and is therefore more relevant to searches in content analysis logic.

The QCA method relies on an iterative process between cases and variables. Data analysis involves feedback on qualitative data and case studies (Legewie, 2013). During this process, it is common to add or delete cases based on combinations of conditions or consequences obtained (Wagemann, 2007).

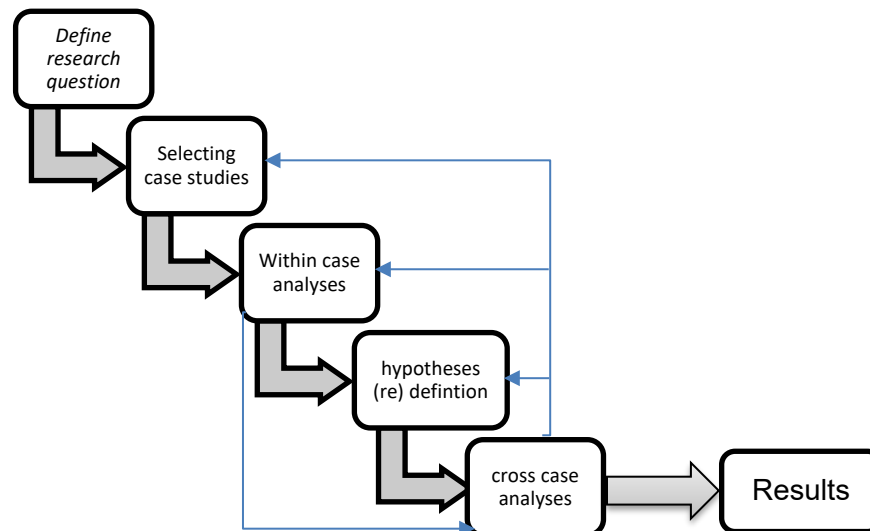


Figure 1: The data analysis process according to the QCA method (Adapted from Legewie, 2013)

Like other methods of quantification of qualitative data (e.g. statistical textual analysis with the ALCESTE method, Artis and Cornée, 2016), this approach requires a very good knowledge of the cases. In fact, unlike quantitative methods, these methods require regular round-trips between the analysis and the data. In the case of ALCESTE, the software does not consider the meaning, literally or figuratively, of the textual forms used. Thus, ALCESTE proposes lines of reflection and interpretation for the researcher to distinguish and interpret classes, and leaves the possibility of not dealing with abnormal phenomena that are difficult to interpret or do not correspond to the hypotheses (Delavigne, 2004). In the QCA method, the treatment of contradictory cases and the study of so-called logical combinations are all stages of interpellation of the researcher and verification of the coherence of the analytical framework. Unlike methods that are not controlled by the researcher in the process, like the black box, the QCA method requires the researcher to master the entire analysis process. The binary coding of the variables requires the researcher to do a detailed job of deconstructing the cases in order to justify them theoretically.

Qualitative case studies are often associated with exploratory research designs (Bergadaà and Nyeck, 1992; Yin, 1989) and quantitative studies on large samples are often aimed at testing a pre-existing theoretical model. The QCA method can be used in both inductive and deductive reasoning. In a deductive reasoning, different conditions assumed to lead to a result are defined, and their impact is observed in the results. In an inductive and exploratory aim, if it is necessary to define the analyzed result, the researcher can propose as many conditions as the data permit, in order to highlight the necessary or sufficient conditions for the results between these variables. The QCA method involves grouping cases into homogeneous categories, defined by specific characteristics. The number of possible categories is 2^k : one criterion defines two categories, two criteria four categories, three criteria eight categories, and so on. As for the number of conditions to be retained, there are mainly two possibilities: either, to limit oneself to a number of categories which is much smaller than the number of cases observed, in order to arrive at a few subsets of the cases, or to include more variables but accept that the observed cases which will be divided into many categories without forming subsets. A deductive research design seems more coherent with a small number of conditions,

whereas in an exploratory view, it is relevant to include a large number of conditions. But it is also possible to explore the data by comparing several "models" with a small number of conditions.

2. APPLYING QCA METHOD TO INTER-ORGANIZATIONAL COLLABORATION

Our research analyzes collaborations between organizations of various natures. Based on a saturation system saturation⁴, we found thirty collaborations, for which the QCA analysis is very relevant. Our research fits well within the logic of content analysis and appropriate for QCA. Using the QCA method, we were able to explore thirty cases of qualitative data and identify success factors for inter-organizational collaboration between big companies and non-profit organizations.

2.1. Case selection and qualitative data collection

We built the case database by identifying renewable energy projects around the world. The data which we have collected is secondary data on the internet. We have included various energy technologies: wind, solar, hydro and geothermal. The data was collected from reports written by international bodies involved in the field (e.g. public investment banks, international renewable energy agencies, and government agencies). We found 30 cases of inter-organizational collaborations within our criteria. Systematically, we identified:

- Stakeholders involved in the project: companies, associations and NGOs, research centers, government agencies, and funders;
- The economic characteristics of the project: location, start date of the project, duration of the project implementation, number of workers, and cost of the project;
- The technological characteristics of the project: technology and quantity of energy produced;
- The characteristics of the country's energy situation: quantity of energy consumed, tension in the energy market, and dependence on a specific source of fossil energy (oil, gas, coal, etc.);
- Acceptance of the project: local people's views on the project (according to the media).

2.2. QCA analysis of the data

We have therefore followed the different steps of QCA methodology (Rihoux and Ragin, 2009): i) selection and coding of the information of the selected cases, ii) construction of the truth table using data dichotomization, iii) verification of the robustness of data (resolution of contradictory cases, Boolean minimization process and taking into account so-called "logical" combinations) and iv) interpretation of the results. We have decided to limit ourselves to three

⁴ We completed our database by adding all the new projects we found that met our criteria (organizational diversity, and renewable energy).

conditions which will create theoretical configurations to eight categories 2^k . Adding an additional condition would have created 16 theoretical configurations, with an average of less than two cases per category. We therefore made a difficult choice because it is simplistic to retain the three conditions that seem to us the most important given the available data:

1. Presence of incentives for renewable energies and innovation: public incentives (tax reductions, dedicated financing, etc.) and incentives from civil society. For example, Singapore has invested more than 570 million US dollars in incentives since 2011, particularly in solar energy solutions (ASEAN Briefing, 2015). Incentives from public policies such as regulations, subsidies, tradable permits, or tax reductions have a positive (direct or indirect) effect on environmental innovation (Depret and Hamdouch, 2009)
2. The presence of a conflict over natural resources: the exploitation of renewable energies can generate conflicts in the management and use of natural resources. Indeed, Ostrom (1990) highlights that renewable resources are more prone to scarcity than abundance, and their exploitation can lead to conflicts between users. Several studies on water management crystallize these conflicts in the possible choices of uses of the resource. These conflicts are also present in land management between alternative uses (agricultural use, energy production or real estate).
3. The international dimension of inter-organizational collaboration. Beyond these elements, the international dimension of collaboration is a major factor contributing to differences between actors (Mohr et al., 2016). Whether at the cultural, linguistic, institutional or political level, the international nature of the collaboration adds to the differences that already exist between the partners in the cases that we have chosen because on one side are the companies and on the other the non-profit organizations.

We have determined whether the projects are successful based on three criteria related to the contribution of the project to the energy transition:

- The amount of energy produced compared to the country's production capacity
- The duration of the project completion, compared to the average time per megawatt required in equivalent technologies.
- The contribution of the project to a sustainable energy supply in the country. This indicator is composed of three variables: the contribution of the project to the diversification of sources of energy production, the impact in terms of environment and reduction of CO² emissions, and the productive capacity of the technology in response to local needs.

We determined the success of the project based on the combination of these indicators (success = 1, failure = 0); only projects rated "1" for at least two of the three criteria were considered successful. This multi-criteria determination of success has the consequence of classifying projects in a situation of failure despite their production of renewable energy. Projects are considered "failures" only if more time is required to complete the project (more than the estimated time initially), energy production is low or environmental impact is less clear. We filled in these factors from the data previously collected, and we coded them in a dichotomous way (table 2) to then build the truth table (table 3).

	Case	Technology	Localization	Incentives	Conflict	International	Success
1	Solar farm in ASYV	Solar	Rwanda	1	0	1	1
2	Hydropower in Intibucá	Hydro	Honduras	0	1	1	0
3	Hydropower Nepal	Hydro	Nepal	1	0	0	1
4	Solar Lantern Rental System Laos	Solar	Laos	0	0	0	1
5	Salkhit wind farm	Eolien	Mongolia	1	0	1	1
6	Sapphire wind farm	Eolien	Pakistan	0	0	1	1
7	Lesedi & Letsatsi projects	PV	South Africa	1	0	1	1
8	Hami project	CPV	China	1	0	1	0
9	Wind farm in Singida	Wind	Tanzania	0	0	1	1
10	Karadzhalovo solar park	Solar	Bulgaria	1	0	1	1
11	Wind farm in Madhya Pradesh, Andhra Pradesh	Wind	India	1	0	0	1
12	Energy efficiency	Energy efficiency	China	1	0	0	0
13	IVANPAH	Solaire Thermal	USA	1	0	0	1
14	Ouarzazate Solar Power Station "Phase one"	Solaire PV	Morocco	1	0	1	1
15	Bugoye Hydro Power	Hydro	Uganda	0	0	1	1
16	The Lake Turkana Wind Power Project (LTWP)	Wind	Kenya	1	1	1	1
17	Palo Viejo Hydro Plant	Hydro	Guatemala	1	1	1	0
18	San Jacinto-Tizate Geothermal Project	Geo-thermal	Nicaragua	1	0	1	1
19	Jasper Solar Photovoltaic Power Plant	Solaire PV	South Africa	1	0	1	1
20	Amanecer Solar CAP Power Plant	Solaire PV	Chile	1	0	1	1
21	Chaglla Hydroelectric Power Plant	Hydro	Peru	0	0	1	1
22	Ulubelu II Geothermal Power Plant	Geo-thermal	Indonesia	1	1	1	0
23	Dai Ninh Hydro Plant	Hydro	Vietnam	0	1	1	0
24	Burgos Wind Project	Wind	Philippines	1	0	1	1
25	Rampur Hydro Project	Hydro	India	1	1	0	0
26	Agua Prieta II	ISCC	Mexico	1	0	1	0
27	Augustin Fresnel 1, Targassonne	Solar	France	1	0	0	0
28	Hassi R'mel	ISCC	Algeria	1	0	1	1
29	Khi Solar One	CSP	South Africa	1	0	1	0
30	Manchasol 1	Th. Solaire	Spain	1	1	1	0

Table 2: Description and coding of the projects

Configuration	Incentives	Conflict	International	Cases	Number of cases	Result (successful or not)
1	0	0	0	4	1	positive
2	1	0	0	3,11,12,13,27	5	Positive with 2 contradictory cases
3	0	1	0	-	0	-
4	1	1	0	25	1	Negative
5	0	0	1	6,8,9,21,15	5	Positive with 1 contradictory case
6	1	0	1	1,5,7,10,14,18,19,20,24,26,28,29	12	Positive with 2 contradictory cases
7	0	1	1	2,23	2	Negative
8	1	1	1	16,17,22,30	4	Negative with 1 contradictory case

Table 3: Truth table of the success factors of the 30 case studies

2.3. CQA Analysis Results

Table 3 presents the eight possible configurations from our three variables; seven configurations can be observed in the data. Configuration number 6 is the most common. It is characterized by international projects that have not had conflict and are located in countries where there are incentives for renewable energies (12 projects). However, we have two contradictory cases in this configuration (cases 26 and 29). Configurations 2 and 5 have five cases each, configuration 5 has one contradictory case, and configuration 2 has two. Configuration 5 groups together cases that are international, have not had conflicts and had no incentive; and configuration 2 brings together local cases, without conflict and with incentives. Configuration 8 is interesting because it groups failed projects, except for one contradictory case. Paradoxically, the simultaneous presence of three factors leads to failed projects. Contradictory cases were re-evaluated from the available data, allowing these contradictions to be resolved. For example, case 8 has been classified as unsuccessful because the duration of implementation of the project is above average (1.7 MW per month against an average time for solar projects of 4 MW / month) and the production is considered low (production of 20.5MW compared to a total electricity production in China of 1223GW). These two criteria led to classification of this case as unsuccessful. However, this project is based on solar technology, while production in China is mainly fossil and hydroelectric. It therefore participates in the diversification of energy sources in a context of increased needs and international pressure (signature of COP21 agreements). In view of these complements, we finally considered this case a success. The same procedure for resolving contradictions was carried out in all configurations.

Four configurations are possible to achieve success. In all these configurations, there is no conflict over natural resources. This criterion is therefore a sufficient condition for a positive result.

The most observed configuration shows that two causal relationships are possible:

- If there is no conflict over the resources, then the project will be a success;
- If the project is international and there are incentives, then it will be a success;

By adopting the norms of Boolean algebra notation (Chanson et al., 2005, Ragin, 1987), the equation is as follows:

$$\text{SUCCES} = \text{INCITATIONS} \cdot \text{INTERNATIONAL} \cdot \text{conflit}$$

This configuration represents 12 cases of our total population, and 12 of the 20 success cases (60%).

However, the homogeneous distribution of cases between configurations 2 and 5 results in imprecision between the "incentive" and "international" factors. Of the 23 cases where incentives were available, nine cases were unsuccessful. Of the 23 projects with actors from different countries, eight failed. Therefore, it seems that incentives are a sufficient condition. However, the differentiation between these two conditions (international and incentive) needs to be further developed. The QCA method makes it possible to highlight the diversity of the possible causal relationships between the conditions and the result. There are then several modalities to arrive at a situation of success.

The consideration of logical cases makes it possible to validate the robustness of the method and the conceptual framework chosen for this study (Carsten and Schneider, 2012). Of the eight possible configurations, a configuration (3) was not observed, it corresponds to the following logical case: a configuration of failure in the presence of conflict over resources, without international dimension or incentive to renewable energies. We conclude that the "absence of conflict" condition leads to a positive result.

Another perspective for future work would be to compare multi-stakeholder collaboration with business-to-business collaboration to challenge the differences between these different types of collaboration.

3. METHODOLOGICAL DISCUSSION

Based on the example of our research, we now return to key aspects of the QCA method and their challenge for researchers using this method: the number of conditions selected, the binary nature of the variables, the causality analysis rather than probabilities, and the generalization of the results obtained.

3.1. The number of integrated variables ("conditions")

One of the limitations of the QCA method is the exponential increase in the number of possible configurations depending on the number of criteria retained. Adding a variable will double the number of configurations, three criteria produce eight possible configurations, four criteria, 16, and so on. Principally, there are two logics to decide the number of conditions: either, the objective is to arrive at categories including several cases and, thus, with a sketch of typology (but it is then necessary to limit the number of conditions); Or, the researcher wishes to explore

the relevance of a maximum of variables even if each category is illustrated only by one or a few cases, or none. If for a small number of variables, a manual analysis is quite simple to perform, the multiplication of variables may justify the use of software. 61% of 313 studies were published between 1984 and 2011 using the QCA method with more than four variables (Rihoux et al., 2013) - but these studies date mostly from 2004, when QCA computer analysis software became widespread.

We chose the first option: with 30 collected cases, we limited ourselves to eight configurations, i.e. three variables or conditions. We could enrich our results by adding an additional variable. This development would allow us to consolidate the importance of the "absence of conflict" factor, to refine the understanding of "international" factors and "incentives" and to explore an additional dimension, for example, the "innovation" dimension of technologies.

3.2. The binary nature of the variables

The binary coding of variables is one of the rules of the method in its original version ("QCA-set QCA"), which can be achieved without reference to a particular software. In our research, we explored a very particular type of inter-organizational collaboration, that is, between at least one company and at least one non-profit organization, dealing with the implementation of a project in the renewable energy sector. Since this exploration was global, we had no choice but to limit ourselves to secondary data collected on the internet. Although we cross-referenced sources and documents, the data at our disposal remain much poorer than those that can be collected *in situ*, through semi-structured interviews and direct observations. This is on the one hand unfortunate - but on the other side, particularly compatible with the QCA method because the coding of data in binary variables depletes the richness of the collected data less than other methods.

The binary encoding of the data represents a limit of the method, because many studied dimensions correspond to continuous variables. In our study, measuring project success and quantifying it was a challenge. The challenge was to not rely only on a quantitative measure related to energy production, and to question inter-organizational collaboration. For example, the use of the "duration of project implementation" criterion was also intended to measure the quality of inter-organizational collaboration. In fact, the forms of hybrid organizations are modes of coordination between agents that are alternatives to the market and to the competition, which raise criticisms and questions about their effectiveness, particularly in terms of time (Ménard, 2003). This aspect of inter-organizational collaboration is therefore an essential dimension but this variable is by definition a continuous variable. It is in response to this limit that the QCA "fuzzy-set" and "multi-value set" variants have been developed, integrating non-binary variables. But this debate goes beyond the opposition of binary variables and continuous variables. Adopting the QCA method is equivalent to entering into a logic of typologies (Fiss, 2011), to question whether or not a case is part of a type, and not to measure intensities. Typologies form a central pillar of strategic management and organizational theory (Fiss, 2011), and some are extremely popular within the discipline. The binarity of well-chosen variables can potentially lead to new ones.

3.3. A causal diversity rather than the probability of a single model

Even before starting an analysis based on Boolean algebra, the intermediate stage of the truth table has, in our opinion, virtues for case study research. It requires researchers to clearly identify the main elements of their theoretical model (emerging or pre-existing), to differentiate possible causes of the consequences in the formulation of their problem, and to systematically question all the cases as to the presence or absence of each of these variables. The presentation as a table (like the truth table of the QCA method) has the advantage of simplicity and clarity. Certainly, such a summary table of results may also be used for presentation purposes only, without necessarily seeking causalities and without giving rise to a subsequent quantitative analysis (for example, Bartel-Radic, 2013). The QCA method thus favors the distance of the researcher from the field to reinforce the objectivity of the analysis. Moreover, in certain situations, this distance from the analysis unit may be an alternative solution to physical or linguistic difficulties of access to the data. Initially presented as an opportunity to analyze small to intermediate samples (Ragin, 1987), QCA is now often used for larger samples. 17% of the 313 studies analyzed by Rihoux et al. (2013) have samples of more than 100 observations, and other methods of statistical analysis (regressions, structural equations, ...) could have been used as well. A majority of the studies involve samples of more than 30 observations that could have been analyzed with the PLS method, for example. More than a default method, QCA is a different way of addressing causality. When statistical models test a single causal model, QCA admits a causal diversity, i.e. the coexistence of several parallel paths to arrive at a result (Chanson et al., 2005). In our study, if the absence of conflicts over natural resources appears to be a sufficient condition, several combinations of configurations lead to successful collaborations. The contradictory cases (we had to treat several in our study), rather than "atypical observations" to put away, represent an opportunity to systematically interrogate the theoretical frameworks and observed cases, allowing the researcher to objectify one's research, questioning all the time one's postulates and hypotheses.

3.4. The generalization of the obtained results

The previous point directly impacts the possibility of generalizing the obtained results. One of the challenges of the QCA method lies in the generalization conditions of the results. The initial goal of the method is to be "case-centered" (Ragin, 1987), which is an advantage but also a disadvantage. "Case dependence" has been repeatedly cited as a limitation of QCA (Rihoux et al., 2014). In the case where the research relates to a whole population of cases exhaustively (one of the initial criteria of the QCA, cf. Chanson et al., 2005), the question of generalization arises only outside this population.

We cannot verify the completeness of our sample, but it is very likely that we have not identified all of the existing renewable energy projects on the planet involving businesses and non-profit organizations - in other words, that the sample is not exhaustive. This lack of data questions the generalization of our results.

As with case study research in general, if statistical generalization is impossible, analytical generalization is feasible and useful (Yin, 1989). It is based on a detailed description of the case context, which must therefore accompany the QCA analysis. In other words, if the "table

of truth" was built following the quantification of qualitative data, the researcher has rich data that allow us to contextualize and specify the results obtained through QCA analysis. On the other hand, if the variables analyzed in the QCA had been collected differently (in pre-existing quantitative databases, for example), this point could be more problematic. The combination of the QCA method with other methods of analysis, qualitative or quantitative, is pointed out as highly desirable by some researchers (Schneider & Wagemann, 2012). However, it is absent from 61.3% of the research referenced by Rihoux and colleagues (2013), in which the QCA analysis does not coexist with any other method of data analysis.

CONCLUSION: RECOMMENDATIONS FOR THE USE OF THE QCA METHOD IN RESEARCH IN MANAGEMENT

The purpose of this contribution was to present and discuss the contributions and limitations of the QCA method for researchers working with qualitative case studies. We chose to explain the steps of the QCA process by systematically questioning its specificities compared to other research methods. We discussed the choices that are necessary for any researcher who wishes to use this method. We conducted our research on inter-organizational collaboration in the field of renewable energies.

Given the limitations of qualitative and quantitative data collection and processing methods, the QCA method offers promising prospects. It establishes causalities between qualitative variables for mid-size case samples. It enriches qualitative and quantitative approaches while ensuring statistical rigor; it retains the richness of qualitative approaches while allowing a generalization approach. This method is also a challenge for the researchers because it forces them to systematically question their theoretical frameworks and their assumptions. The control of the data (of their nature, collection and processing) and the explanation of the different stages of the research design are essential to the good use of this method. However, and despite its recent strong development, this method remains confined to a few circles of researchers initiated to debates on the epistemology of methods in the social sciences.

The use of QCA has given us an important result: inter-organizational collaboration in renewable energies is a success in contexts where there is no conflict over the management and use of natural resources. The new approaches to natural resource management developed by Elinor Ostrom and his collaborators show that conflict is ubiquitous in the collective management of common resources. In this context, the factory of institutions is in itself "a difficult, time-consuming process, a source of conflict" (Ostrom, 1990: 14). It theorizes conflict as a cognitive problem due to differences in interpretation, and thus presents the establishment of institutions within the community as a process allowing interaction and interdependence between actors, which can increase mutual trust. and decrease conflicts. One of the perspectives for our research would then be to study the Ostrom institutions set up during these successful collaborations. In other words, how has collaboration with non-profit organizations helped to avoid major conflicts over resources and thus facilitate the faster implementation of more significant projects involving renewable energies?

As we have discussed in this chapter, the generalization of this result is not as obvious from our data as from the method used. The choice of a limited number of variables and the binary

coding greatly simplify the results. But the QCA method also allowed us to discuss the different configurations and their combination to achieve a successful collaboration on a vast field of exploration. Even though we were not able to carry out *in situ* observations directly, the QCA method enriched the debate on this topic.

REFERENCES

- Allard-Poesi F. (2003), « Coder les données », in : Giordano Y. (dir.) *Conduire un projet de recherche-Une perspective qualitative*, EMS-Management et société, p. 245-290.
- Artis A. et Cornée S. (2016), « La face cachée de l'intermédiation financière : composition, traduction et mémorisation du savoir idiosyncratique dans la banque solidaire », *Systèmes d'Information et Management*, vol. 21, n° 3.
- ASEAN Briefing (2015), « Incentives For Renewable Energy Investment », Singapore, N.p.
- Barney J. (2001), « Resource-based theories of competitive advantage: A ten-year retrospective on the resource-based view », *Journal of Management*, vol. 27, n° 6, p. 643-650.
- Bartel-Radic A. (2013), « 'Estrangeirismo' and Flexibility: Intercultural Learning in Brazilian MNCs », *Management International*, vol. 17, n° 4, p. 239-253.
- Bergadaà M. et Nyeck, S. (1992), « Recherche en marketing: un état des controverses », *Recherche et Applications en Marketing*, vol. 7, n° 3, p. 23-44.
- Bianchi P. et Labory S. (2013), « Structural Transformations in Industry and Filières », *Revue d'économie industrielle*, n° 144, p. 177-199.
- Chanson G., Demil B., Lecocq X. et Sprimont P. A. (2005), « La place de l'analyse qualitative comparée en sciences de gestion », *Finance Contrôle Stratégie*, vol. 8, n° 3, p. 29-50.
- Chanteau J.-P. (2011), « L'économie de la RSE : éléments de méthode institutionnaliste », *Revue de la régulation*, n° 9.
- Curchod C. (2003), « La méthode comparative en sciences de gestion : vers une approche quasi-expérimentale de la réalité managériale », *Finance Contrôle Stratégie*, vol. 6, n° 2, p. 155-177.
- Davis J. P. (2016), « The Group Dynamics of Inter-Organizational Relationships: Collaborating with Multiple Partners in Innovation Ecosystems », *Administrative Science Quarterly*, vol. 61, n° 4, p. 621-661.
- Delavigne V. (2004), « Présentation d'Alceste », *Revue Texto!*, n°9, p. 324-329.
- DeMeur G. et Rihoux B. (2002), *L'Analyse Quali-Quantitative Comparée*, Academia Bruylant.
- Den Hond F., de Bakker F. G. et Doh J. (2015). « What prompts companies to collaboration with NGOs? Recent evidence from the Netherlands ». *Business & Society*, vol. 54, n° 2, p. 187-228.
- Depret M.-H. et Hamdouch A. (2009), « Quelles politiques de l'innovation et de l'environnement pour quelle dynamique d'innovation environnementale ? », *Innovations*, n° 29, p. 127-147.
- Dütting G. et Sogge D. (2010), « Building Safety Nets in the Global Politic: NGO collaboration for solidarity and sustainability », *Development*, vol. 53, n° 3, p. 350-355.
- Dyer J. et Singh H. (1998), « The Relational View: Cooperative Strategy and Sources of Interorganizational Competitive Advantage », *Academy of Management Review*, vol. 23, n° 4, p. 660-679.
- Eisenhardt K. M. (1989), « Building Theories from Case Study Research », *Academy of Management Review*, vol. 14, n° 4, p. 532-550.

- Elyachar J. (2003), « Mappings of power: the state, NGOs, and international organizations in the informal economy of Cairo », *Comparative Studies in Society and History*, vol. 45, n° 3, p. 571–605.
- Elyachar J. (2005), *Markets of dispossession: NGOs, economic development, and the state in Cairo*, Duke University Press Books.
- Fiss, P. C. (2011), « Building Better Causal Theories: A Fuzzy-Set Approach to Typologies in Organization Research », *Academy of Management Journal*, vol. 54, n° 2, p. 393-420.
- Forest J. et Hamdouch A. (2009), « Les clusters à l'ère de la mondialisation : fondements et perspectives de recherche », *Revue d'économie industrielle*, n° 128, p. 9-20.
- Gioia D. A., Corley K. G. et Hamilton A. (2012), « Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology », *Organizational Research Methods*, vol. 16, n° 1, p. 15-31.
- Grawitz M. (2001), *Méthodes des sciences sociales*, Dalloz.
- Gray B. (1989), « Developing a Theory for Collaboration », in: Gray B. (dir.), *Collaborating: Finding common ground for multiparty problems*, New Jersey, Jossey-Bass, p. 226–245.
- Greckhamer T., Misangyi V. F., Elms H. et Lacey R. (2008), « Using qualitative comparative analysis in strategic management research: An examination of combinations of industry, corporate, and business-unit effects », *Organizational Research Methods*, vol. 11, n° 4, p. 695-726.
- Haesebrouck T. (2016) « The Added Value of Multi-Value Qualitative Comparative Analysis », *Forum: Qualitative Social Research*, vol. 17, n° 1, art. 12, p. 1-29.
- Klein J. L., Laville J. L. et Moulaert, F. (2014), *L'innovation sociale*, Erès.
- Laurent C. et Du Tertre C. (dir.) (2008), *Secteurs et territoires dans les régulations émergentes*, Paris, L'Harmattan.
- Legewie N. (2013), « An introduction to applied data analysis with qualitative comparative analysis », *Forum: Qualitative Social Research*, vol. 14, n° 3, p. 1-45.
- Maurer B. (2012), « Mobile money: Communication, consumption and change in the payments space », *Journal of Development Studies*, vol. 48, n° 5, p.589–604.
- Ménard C. (2003), « Économie néo-institutionnelle et politique de la concurrence les cas des formes organisationnelles hybrides », *Économie rurale*, n°277-278, p. 45-60.
- Mohr A. T., Wang C. et Goerzen A. (2016), « The impact of partner diversity within multiparty international joint ventures », *International Business Review*, vol. 25, n° 4, p. 883-894.
- Ostrom E. (1990), *Governing the Commons: The evolution of institutions for collective action*, Cambridge University Press.
- Ragin C.C. (1987), *The Comparative Method*, University of California Press.
- Ragin C.C. (2000), *Fuzzy-Set Social Science*, University of Chicago Press.
- Ren H., Gray B. et Kim, K. (2009), « Performance of International Joint Ventures: What Factors Really Make a Difference and How? », *Journal of Management*, vol. 35, n° 3, p. 805-832.
- Reverdy T. (2014), *La construction politique du prix de l'énergie: sociologie d'une réforme libérale*, Presses de Sciences Po.
- Rihoux B., Marx A. et Alamos-Concha P. (2014), « 25 années de QCA (Qualitative Comparative Analysis) : quel chemin parcouru ? », *Revue internationale de politique comparée*, vol. 21, n° 2, p. 61-79.
- Rihoux B., Alamos-Concha P., Bol D., Marx A. et Rezsöhazy I. (2013), « From niche to mainstream method? A comprehensive mapping of QCA applications in journal articles from 1984 to 2011 », *Political Research Quarterly*, vol. 66, n° 1, p. 175-184.

- Rihoux B. et Ragin C. (2009), *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*, New York, Sage.
- Schneider C. Q. et Wagemann C. (2010), « Standards of good practice in qualitative comparative analysis (QCA) and fuzzy-sets », *Comparative Sociology*, vol. 9, n° 3, p. 397-418.
- Schneider C. Q. et Wagemann C. (2012), *Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis*, Cambridge, Cambridge University Press.
- Seny Kan A. K., Adegbite E., El Omari S. et Abdellatif M. (2016), « On the use of qualitative comparative analysis in management », *Journal of Business Research*, vol. 69, n° 4, p. 1458-1463.
- Sims C. A. (2010), « But Economics Is Not an Experimental Science », *Journal of Economic Perspectives*, vol. 24, n° 2, p. 59-68.
- Smith N. C. (2003), « Corporate social responsibility: Whether or how? », *California management review*, vol. 45, n° 4, p. 52-76.
- Thomas J., O'Mara-Eves A. et Brunton G. (2014), « Using qualitative comparative analysis (QCA) in systematic reviews of complex interventions: a worked example », *Systematic reviews*, vol. 3, n° 1, p. 679-706.
- Wagemann C., Buche J., et Siewert M. B. (2016), « QCA and business research: Work in progress or a consolidated agenda? », *Journal of Business Research*, vol. 69, n° 7, p. 2531-2540.
- Yin R. (1989), *Case Study Research: Design and Methods*, Sage.

AUTHORS

Amélie Artis

Associate Professor in Economics
Univ. Grenoble Alpes, Sciences Po Grenoble, CNRS, PACTE, F-38000 Grenoble, France

Anne Bartel-Radic

Full Professor in International Management
Univ. Grenoble Alpes, Grenoble INP, CERAG, F-38000 Grenoble, France

Hezam Haidar

PhD Candidate in Industrial Engineering
Univ. Grenoble Alpes, Grenoble INP, CNRS, G-SCOP, F-38000 Grenoble, France

Les méthodes expérimentales dans la recherche en management international

RÉSUMÉ

Les méthodes expérimentales sont depuis longtemps employées en Sciences naturelles afin de tester une hypothèse en contrôlant les conditions d'expérimentation. Ces méthodes ont été reprises en Sciences économiques et de gestion ces dernières décennies, et sont devenues une référence dans certaines sous-disciplines comme le marketing. Or, les expérimentations restent rares dans la recherche en management international. Dans ce chapitre, nous présentons les difficultés, mais aussi les avantages des méthodologies expérimentales dans cette sous-discipline à travers l'exemple du projet InterCCom, portant sur la compétence interculturelle. Grâce au développement d'un jeu sérieux informatisé dans le cadre de ce projet, les comportements des membres d'une équipe internationale virtuelle sont mesurés, ce qui permettra de répondre à des problématiques de recherche variées.

MOTS CLÉS : méthode expérimentale, expérimentation, management international, compétence interculturelle

INTRODUCTION

L'objectif de ce chapitre réside dans le fait de proposer aux chercheurs et aux étudiants en sciences sociales une introduction aux méthodologies expérimentales. Ces méthodologies proviennent à l'origine des sciences naturelles comme la physique, la chimie ou la médecine. Lors des dernières décennies, elles ont largement été reprises en économie et en Sciences de gestion, notamment dans la sous-discipline du marketing. En revanche, en management international, les contributions restent relativement rares, alors que les meilleures revues de la discipline ont appelé de leurs vœux une forte augmentation de ces recherches.

Ce chapitre est structuré en quatre parties :

1. L'expérimentation clinique telle qu'elle est utilisée en médecine ou en neurosciences.
2. L'utilisation des méthodes expérimentales en économie et en sciences de gestion depuis plusieurs décennies.
3. L'application au champ du management international et les difficultés inhérentes à cette sous-discipline.
4. La présentation d'un projet de recherche en management interculturel et la manière dont il mobilise les méthodes expérimentales.

1. LES MÉTHODES EXPÉRIMENTALES : QUELQUES BASES

Au sens large, on peut qualifier d'expérimentation toute collecte de données empiriques qui vise à tester une hypothèse, et ce quel que soit le mode de collecte de données (observation, interviews, focus group, ou encore questionnaire). Les hypothèses à tester portent notamment sur des liens entre une ou plusieurs causes et une conséquence. Un lien entre deux phénomènes peut être qualifié de causal si les critères suivants sont remplis (Hill, 1965) :

- la corrélation doit être forte ;
- cette corrélation doit se montrer dans différents contextes, dans des études menées sur différentes populations ;
- la cause devrait avoir des conséquences spécifiques ;
- la cause devrait précéder la conséquence dans le temps ;
- il doit être possible de trouver une explication plausible de la manière dont la cause a influencé la conséquence ;
- et en général, il doit être possible de trouver une fonction monotone entre la cause et la conséquence (« plus » de cause devrait entraîner « plus » de conséquence).

Au sein des méthodes expérimentales, on distingue les « vrais » designs expérimentaux de la quasi-expérimentation et d'autres techniques de collecte de données. Selon Greenwood (1972 : 178), les vrais designs expérimentaux doivent remplir les critères suivants :

- ils visent à tester une hypothèse ;
- ce test passe par un protocole de recherche précis et maîtrisé ;
- et les conditions d'expérimentation sont contrôlées à l'aide de techniques d'élimination de variables, de conditions constantes, et d'échantillonnage randomisé.

Seuls les designs de recherche comprenant deux mesures, avant et après un événement ou un traitement proposé au groupe testé (par exemple l'inclusion d'une variable), et avec un groupe de contrôle n'ayant pas subi le même traitement, remplissent entièrement ces conditions. On pensera ainsi, en pharmacologie, à un groupe de malades souffrant d'une pathologie dont certains se sont vus administrer un médicament, alors qu'une autre partie du groupe a reçu un placebo, et une évaluation de leur état de santé avant et après le traitement. On distingue donc les protocoles expérimentaux qui se déroulent dans l'environnement habituel des personnes (« field experiment ») des protocoles expérimentaux qui se déroulent dans un lieu dédié (en « laboratoire » ou « clinique »). Les études se déroulant dans l'environnement habituel des personnes ont une validité externe plus élevée car les comportements des personnes sont moins « dénaturés » par l'expérimentation. Mais non seulement ces déplacements des chercheurs sur le terrain sont plus coûteux en temps et en moyens, mais en plus elle ne permet pas non plus de contrôler les variables contextuelles avec la même précision. On confère ainsi habituellement aux expérimentations cliniques une plus forte validité interne, même si celle-ci rencontre également des limites (Schnell, Hill & Esser, 2018).

Le fondement même de l'expérimentation clinique repose sur trois piliers incontournables :

1. Une reproduction « factice » qui soit aussi fidèle que possible de l'environnement dans lequel l'expérimentation prendra place.
2. Un choix judicieux des outils et des méthodes à mobiliser pour parvenir à ses objectifs.
3. Une question de recherche à la fois claire, mesurable et pertinente.

Pour que le premier pilier soit stable, une bonne technique consiste à longuement observer la réalité. Ainsi, en termes de management interculturel, une immersion dans la culture du pays, un partage fréquent avec ses représentants et un travail précis sur les stéréotypes peuvent s'avérer être très efficaces. Pour que le deuxième pilier soit stable, il suffit de justifier chaque décision ; on devrait avoir le terme « parce que » en permanence à l'esprit. Quant au troisième pilier, une seule question mérite récurrence : que veut-on mesurer ?

On qualifie de designs quasi-expérimentaux des designs de recherche dans lesquels on compare bien un groupe expérimental et un groupe de contrôle, mais où le classement d'un individu dans un groupe ou dans l'autre n'est pas fait de manière randomisée. Généralement, ce sont les individus observés qui choisissent eux-mêmes le groupe auxquels ils veulent appartenir, ce qui crée des biais. Au sein de ces designs quasi-expérimentaux, on distingue les « expérimentations naturelles » et les « quasi-expérimentations ». Dans les premières, la manipulation de la variable intervient sans action des chercheurs (comme par exemple la décision de réduire la vitesse sur les routes nationales de 90 km/h à 80 km/h en France en 2018), alors que dans les secondes, ce sont les chercheurs qui manipulent délibérément la variable (comme l'introduction d'une nouvelle politique publique uniquement sur une ou quelques zones géographiques expressément choisies pour vérifier son impact).

L'utilisation des méthodes expérimentales en sciences sociales se heurte à un certain nombre de limites. Il faut de très grands échantillons (par exemple 1500 personnes si l'on souhaite pouvoir généraliser les résultats à la population d'un pays comme l'Allemagne ; Schnell, Hill & Esser, 2018 : 203), personnes idéalement choisis au hasard, ce qui nécessite des ressources extrêmement importantes. Mais surtout, de nombreuses variables étudiées en sciences sociales, comme la culture, l'intelligence, ou les catégories socio-professionnelles ne peuvent pas facilement être manipulées durant l'expérimentation et surtout ne permettent pas d'attribuer les personnes au hasard à l'un ou l'autre des deux groupes.

Prenons le cas d'une mesure de l'intelligence. Le facteur général (ou facteur g - Spearman, 1904) consistera à s'accorder sur une conceptualisation de l'objet de recherche, à la fois dans sa définition et dans sa structure récurrente afin de pouvoir rapidement identifier toute dissonance. Ici, il est essentiel de distinguer le sujet capable de résoudre un problème parce qu'il l'a déjà résolu, parce qu'il a observé sa solution ou bien parce qu'il fait appel à son unique intelligence. Ainsi, on distingue la mémoire, l'apprentissage et l'intelligence « pure » (Mouillot, Drillon, & Montargot, 2018). Dans une recherche sur l'intelligence, nous pourrions déplorer que les seuls tests de facteur g encore conçus de nos jours soient des matrices. Or les matrices n'éliminent pas le culturel : elles mettent en œuvre de nombreuses dimensions cognitives (ce qui fait leur force), parmi lesquelles, par exemple, la mémoire de travail et la verbalisation. Et puis il est incontestable qu'une telle approche ramène à la psychologie

différentielle. Celle-ci fait de son objet d'étude l'existence chez les individus des caractéristiques stables qui les distinguent les uns des autres.

Créer une expérimentation clinique permet donc non seulement de faire apparaître les différences, y compris sur des dimensions moins apparentes que la couleur des cheveux, mais encore de mesurer cette stabilité, car la stabilité des différences fonde le pronostic. Avant toute création d'un outil d'expérimentation clinique, il est donc essentiel de s'accorder sur le cœur et la limite de ce que l'on souhaite mesurer. Si, dans le cadre de l'intelligence, par exemple, les chercheurs décident que l'intelligence est la capacité à résoudre un problème encore jamais rencontré, alors il leur sera possible de faire varier les éléments cliniques du test selon, autre exemple, des degrés de difficulté, le facteur g de la nouveauté du problème à résoudre étant systématiquement cliniquement présente.

Venons-en au test. Cette technique est généralement utilisée pour établir des différenciations de manière numérique, notamment grâce à des outils statistiques de type dispersion, écart-type, etc. Mais il est également possible d'utiliser le test afin de mettre le sujet dans une situation expérimentale atypique néanmoins efficace. Prenons le cas du *Thematic Aperception Test* (T.A.T.) créé en 1935 par Henry Murray. Fréquemment utilisé en psychologie clinique en parallèle du Test d'Herman Rorschach, le T.A.T. consiste en une série de planches présentant des scènes de la vie quotidienne. A la différence du Rorschach où le patient est sollicité pour savoir ce qu'il « voit » dans les « tâches » qui lui sont présentées, dans le T.A.T. le patient est sollicité pour raconter une histoire autour de ce qu'il voit. Il fait donc appel à une multitude d'éléments complexes et subconscients, éléments qui le guident dans son quotidien mais qui, en l'occurrence, sont mobilisés artificiellement par le biais d'une suggestion clinique. Il devient alors très intéressant d'utiliser des variantes de ce type de test afin de connaître le ressenti intime d'un sujet. Ainsi, par exemple, plutôt que de lui demander s'il pense qu'une Mercedes est une marque jeune ou plutôt mature, ou encore si le lait fait grossir, il suffit de lui présenter l'image de deux silhouettes, l'une fine, l'autre plus enrobée et de lui demander laquelle boit du lait ou encore laquelle conduit une Mercedes. Les associations subconscientes feront le reste. Enfin, l'analyse factorielle permet, à partir des résultats d'un test, d'isoler des facteurs communs à plusieurs tests afin de permettre l'interprétation de constats numériques ou qualitatifs. Dans ce cas, nous sommes vraiment au cœur de l'analyse clinique comparative au travers du groupe de contrôle face aux groupes expérimentaux. L'analyse factorielle permet de clairement identifier les variantes et leur impact sur les sujets appartenant aux groupes de contrôle. Par exemple, si l'on souhaite mettre en place un protocole permettant de suggérer artificiellement et subconsciemment à des sujets de fournir certaines réponses à un questionnaire, il est primordial de commencer par soumettre le questionnaire à un groupe de contrôle afin, ensuite, d'être en mesure de pointer les variations.

2. L'EXPÉRIMENTATION EN ÉCONOMIE ET GESTION

En 1993, Davis et Holt écrivaient que seulement peu de chercheurs en économie utilisaient des expérimentations en laboratoire pour évaluer des propositions économiques sous des conditions contrôlées attentivement, mais que leur nombre était croissant. La situation a radicalement évolué depuis l'époque, et l'économie comportementale est devenu un vaste champ de recherche, largement reconnu. Elles sont utilisées pour tester les principales

théories économiques, comme par exemple la prise de décisions d'investissement, la disponibilité des biens publics, l'équilibre des marchés, les politiques publiques et d'autres variations en économie environnementale (Jacquemot, l'Haridon & Morin, 2013 ; Lunn & Choidealbha, 2018 ; Ferraro & Price, 2013 ; Binet, Denant-Boemont & Hammiche, 2019). En économie, l'expérimentation connaît un succès croissant car les expériences en laboratoire ou dans l'environnement habituel des personnes permettent aux chercheurs d'observer et d'analyser les déterminants spécifiques du comportement économique (Jacquemot, l'Haridon & Morin, 2013). Harrison et List (2014) précisent qu'une expérience en laboratoire est une méthode efficace pour créer des scénarios factices. Lunn et Choidealbha (2018) pensent qu'une étude bien conçue en laboratoire peut être la meilleure méthode pour répondre aux questions de recherche posées par le Législateur.

En Sciences de gestion, le recours aux méthodes expérimentales s'est également renforcé (voir Igalens & Roussillon Soyer, 2019), mais plus tardivement, et dans une moindre mesure. Le marketing et la finance comportementale sont les champs de recherche où elles sont les plus fréquentes. Ici aussi on trouve des problématiques de recherche pouvant relever de choix de consommation ou d'achat (de titres par exemple).

Des designs de recherche qui ne prévoient de mesure qu'à un moment t , et donc pas de mesure avant / après, sont qualifiés de pré-expérimentations. Une méthode de collecte de données empiriques très courante en Sciences de gestion est celle des questionnaires. Dans ces « ex-post-facto-designs », le chercheur peut certes étudier des corrélations, mais pas manipuler la variable indépendante, ce qui ne permet pas autant que l'expérimentation d'éliminer un certain nombre d'explications alternatives.

L'expérimentation clinique consiste au contraire à reproduire une situation réelle dans un cadre ou un environnement artificiel avec le maximum de fidélité pour la réalité dont il est issu afin de tester une variable ou de mesurer une variation (Δ). Pour cela, le chercheur partira soit de données statistiques réelles recueillies à partir d'observations longitudinales, soit il observera le comportement d'un groupe dit de « contrôle » afin d'avoir une base comparative avant d'observer des groupes dits « expérimentaux », donc modifiés de la variable qu'il souhaite mesurer. Par exemple, à partir de statistiques de consommation, le chercheur pourra recréer un environnement proche d'un supermarché et fournir aux sujets une liste de courses avant d'observer la manière dont ils s'orienteront dans les rayons, leurs hésitations, etc. Autre exemple, si l'on souhaite mesurer la cohérence de l'architecture d'un site internet, il suffit de placer des sujets devant un écran et de leur demander de trouver une information avant d'observer le chemin qu'ils emprunteront, cette fois virtuellement.

L'expérimentation clinique ne peut étudier les lois habituelles d'un comportement qu'en faisant varier systématiquement les facteurs ayant une influence sur lui. Or, certains de ces facteurs ne peuvent être modifiés qu'en utilisant les différences existantes entre les individus et en comparant les réponses données par un certain groupe de sujets aux réponses données par un autre correctement choisi, c'est-à-dire à la composition quantitative et qualitative équivalente. Il est donc fondamental d'identifier les facteurs prégnants étant à l'origine du comportement étudié afin soit de les modifier, soit d'en retirer un ou plusieurs, voire d'en ajouter un ou plusieurs. Voici la raison pour laquelle le recours à des groupes de contrôle et expérimentaux est tellement importante, ces groupes devant être similaires dans leurs

représentations et leurs contenus. Sans référence systémique, stable et récurrente, impossible d'observer une saillance, une digression ou une anomalie.

L'exemple suivant est tiré de la recherche doctorale de l'un des auteurs en neuromarketing (Mouillot, 1999). L'objectif était de mesurer le degré de suggestibilité de deux groupes expérimentaux face à une suggestion subliminale, autrement dit subconsciente. Pourquoi deux groupes ? Car il était nécessaire de vérifier les résultats du premier avec une seconde expérience. En effet, si le groupe de contrôle sert de référence avant que la variable clinique ne soit mobilisée, les résultats issus d'un groupe expérimental doivent être confirmés par une seconde expérimentation organisée dans les mêmes circonstances mais avec des sujets différents afin de mettre en exergue toute variation du résultat. Si les deux groupes expérimentaux produisent des données dont la proximité est jugée acceptable - cet intervalle doit être défini en amont de l'expérience - alors l'influence de la variable clinique devient comparable au groupe de contrôle. En revanche, si les résultats fournis par les deux groupes expérimentaux soient divergents franchement, soit convergent mais de manière jugée comme étant non-significative, alors le recours à un troisième groupe expérimental est nécessaire pour identifier le ou les éléments qui sont à l'origine de cette divergence.

Dans l'exemple qui nous préoccupe, nous étions donc face à l'hypothèse selon laquelle un comportement naturel pouvait être influencé subconsciemment et être ainsi modifié pour au moins 30% des sujets soumis à l'expérimentation, et ce à plus ou moins 5 points. Pour cela, nous avons utilisé un groupe de contrôle constitué de 159 sujets et deux groupes expérimentaux respectivement constitués de 95 et de 58 sujets. Méthodologiquement, le nombre de sujets de contrôle était suffisamment proche du nombre de sujets expérimentaux (159 vs. 153) pour autoriser une comparaison. Un questionnaire fut soumis au groupe de contrôle (159 sujets) afin de connaître le degré de pénétration de marques de boissons non-alcoolisées hors suggestion subconsciente et indépendamment de toute suggestion sensorielle. Les sujets ont rempli le questionnaire dans une salle dans laquelle il n'y avait aucune référence à une quelconque marque. Sans surprise, les résultats ont placé - dans l'ordre - Coca-Cola, Evian, Orangina, Perrier et Pepsi-Cola en tête des occurrences. Ensuite, un premier groupe expérimental fut mobilisé (95 sujets). Ce groupe fut soumis à trois clips vidéo dans lesquelles la suggestion de citer la marque de sirop à l'eau Teisseire fut insérée et projetée à une vitesse de 1/25^e de seconde. Le groupe fut invité à regarder ces clips afin de choisir lequel serait le plus pertinent pour accompagner une campagne publicitaire visant à lutter contre le SIDA. Si les mêmes marques furent citées par le groupe expérimental n°1 que par le groupe de contrôle, la marque Teisseire apparut en 4^e position et fit sortir Pepsi-Cola du classement. 33,68% des réponses ayant été modifiées, notre hypothèse était validée. Il ne restait plus qu'à vérifier avec un second groupe expérimental que ces résultats furent méthodologiquement acceptables. Sept mois plus tard, un second groupe expérimental fut donc constitué (58 sujets) et soumis au même protocole. À l'issue du visionnage des clips, la marque Teisseire apparut en 5^e position dans les réponses au questionnaire, remplaçant cette fois-ci la marque Perrier. Dans ce second cas, 27,33% des réponses furent modifiées. Nous avons émis l'hypothèse d'une modification comportementale artificielle comprise dans un intervalle de 25 à 35% et nos deux résultats se situaient dans cet intervalle. La marque Teisseire n'ayant jamais été citée par les sujets du groupe de contrôle et un test de Khi² ayant validé nos statistiques, ce résultat nous encouragea à penser que la suggestion artificielle

subconsciente avait un réel impact sur des sujets choisis de manière aléatoire, hommes comme femmes, et dont l'âge était compris entre 18 et 25 ans.

3. L'EXPÉRIMENTATION EN MANAGEMENT INTERNATIONAL

Les designs de recherche expérimentaux sont de plus en plus fréquemment utilisés en économie, en psychologie ou en marketing, par exemple, où ils ont tendance à devenir obligatoires pour les publications de très haut niveau. En revanche, en management international, Zellmer-Bruhn, Caligiuri et Thomas (2016) ont constaté que moins de 1% de plus de 900 études empiriques publiées dans le Journal of International Business Studies (JIBS) utilisent des méthodes expérimentales. Cela est entre autres dû à la difficulté de concevoir des protocoles de recherche appropriés qui peuvent être affectés de façon aléatoire pour les participants et aux défis de l'échantillonnage.

Jusqu'en 2016, l'expérimentation était très peu présente en management international (Zellmer-Bruhn, Caligiuri & Thomas, 2016). Les rares études identifiées portaient sur le comportement du consommateur, la communication commerciale, les choix d'investissement, les différences culturelles dans la prise de décision, et *l'empowerment* et la satisfaction au travail. Parmi ces recherches, la plupart ne reposent pas sur une vraie expérimentation, car les échantillons ne sont pas définis de manière randomisée.

Dans le contexte du management international, l'échantillonnage randomisée représente une difficulté particulière. Il est souvent impossible de satisfaire ce critère. Comme le soulignent Zellmer-Bruhn, Caligiuri et Thomas (2016), les chercheurs ne peuvent pas simplement attribuer un pays à une économie politique, des entreprises à des stratégies de globalisation, ou un pays d'origine à un individu, pour ne donner que quelques exemples. De plus, les études réalisées auprès d'étudiants sont considérées dans le champ du management international comme d'un niveau moindre, alors qu'elles sont courantes dans des disciplines comme la psychologie. Les échantillons d'étudiants peuvent être appropriés pour certaines études en management international, mais pour certains sujets, des étudiants n'ont pas suffisamment d'expérience et de connaissance pour pouvoir participer à l'expérimentation (Van Witteloostuijn, 2015). Par ailleurs, les recherches en management internationales requièrent parfois des contextes culturels et institutionnels spécifiques et variables, souvent avec des personnes venant de différents endroits (Zellmer-Bruhn, Caligiuri & Thomas, 2016). Ceci représente un défi supplémentaire dans le recrutement des participants à l'expérimentation.

Dans la recherche en management international, les expérimentations les plus courantes portent sur le niveau d'analyse individuel, ou de l'équipe. En marketing international, des expérimentations sont menées depuis plus longtemps. L'étude de Pornpitakpan (1999) en est un exemple intéressant. L'auteur conclut que les Américains qui s'adaptent à deux cultures et langues étrangères, ici le Thaï et le Japonais, réalisent des performances commerciales plus élevées, même dans leur contexte national. Roth, Prasnikar, Okuno-Fujiwara et Zamir (1991) utilisent un design de recherche expérimental pour comparer des comportements de négociation entre plusieurs cultures. Ils montrent que la définition d'une offre « acceptable » dépend de la culture nationale. En management, Caligiuri et Phillips (2003) conduisent une expérimentation qui attribue des participants de manière randomisée à l'utilisation ou non d'un

outil de décision auto-évalué. Ils montrent que les participants qui ont reçu l'outil en même temps qu'une prévision réaliste du travail à accomplir ont un niveau de confiance plus élevé en leur décision d'accepter une expatriation, et ont obtenu de meilleures performances durant cette expatriation.

En management *interculturel* (c'est-à-dire si l'on veut dépasser la comparaison de différentes cultures), l'expérimentation clinique est traditionnellement interactionnelle. En effet, l'interculturalité impliquant nécessairement l'existence d'une relation entre au-moins deux personnes aux référentiels personnels, sociaux et sociétaux différents, cette approche vise à démontrer l'existence d'interactions entre deux sujets. Toute réponse étant considérée comme étant une réaction à un ou plusieurs stimuli, le comportement d'une personne ne saurait être modifié sans modifications concomitantes de son entourage.

Or, en matière d'expérimentation clinique, tout est question d'interprétation, notamment lorsque l'on s'attelle au vaste domaine de l'analyse interculturelle. Sans interprétation, toute mesure est vide de sens puisqu'elle n'est ni relativisée, ni replacée dans son contexte, en l'occurrence expérimental. Ainsi, lorsque l'on parle de la méthode des tests, sans analyse de l'écart-type, une moyenne pourrait apparaître révélatrice d'une tendance alors qu'il n'en est rien. Il en va de même avec la dispersion. A partir d'un même nuage de points sur lequel une grande concentration apparaîtrait séparée de quelques points isolés, une étude des comportements de consommation serait tentée de se focaliser sur le nuage de points alors qu'une étude psychologique ou interculturelle serait inspirée d'observer les points isolés, par exemple respectivement représentatifs de déviations psychologiques ou de spécificités culturelles. Dans une recherche, la nature même des informations qu'il convient de recueillir pour atteindre l'objectif commande donc les moyens employés pour le faire. Ainsi, Madeleine Grawitz nous invite dès 1993 à réfléchir au fait que s'il est possible d'envisager attraper des poissons avec un filet à papillons, on n'attrape certainement pas des papillons avec des hameçons. Friedrich Nietzsche a d'ailleurs toujours soutenu que les richesses les plus précieuses étaient les méthodes, un discours que ne reniera pas René Descartes. Car ce n'est pas l'objet qui fait la science mais bien la méthode.

Malgré le faible nombre d'études mobilisant des méthodologies expérimentales en management international, de nombreux auteurs mettent en avant leur importance future. Zellmer-Bruhn, Caligiuri et Thomas (2016) considèrent que les recherches expérimentales sont insuffisamment présentes dans *Journal of International Business Studies*, mais qu'elles représentent une opportunité à tester des relations causales en management international. Van Witteloostuijn (2015) encourage le développement d'une « tradition expérimentale en management international » à travers la création d'outils de recherche digitaux. Ces outils permettraient par ailleurs d'augmenter l'efficacité et le plaisir des enseignements en management international. Nous présentons dans la section suivante le projet InterCCom, qui vise le développement d'un tel outil utile à la recherche et à la pédagogie, et qui porte sur le travail en équipe interculturelle.

4. LE PROJET INTERCCOM : UNE PLATEFORME POUR L'EXPÉRIMENTATION EN MANAGEMENT INTERNATIONAL

Dans cette dernière partie du chapitre, nous présentons un projet mené depuis janvier 2019 à Sciences Po Grenoble et l'Université Grenoble Alpes, et qui vise à construire un jeu sérieux en management interculturel, appelé *LINK the serious game*®. Il peut être utilisé comme outil de quasi-expérimentation en recherche, car les choix faits par les joueurs sont enregistrés. Le jeu est implémenté par une plateforme digitale également développée au sein du projet, la plateforme GenaGame®. Sur cette plateforme, des contenus (documents, images, vidéos) peuvent être téléversés, et un cheminement interactif du joueur à travers des contenus, en fonction d'un scénario et de dimensions théoriques sous-jacentes, peut être défini. D'autres jeux en management international sont ainsi en cours de développement. L'ensemble de ces jeux sont regroupés sous la marque i-Team Games®.

La pertinence du projet InterCCom au sein du champ du management interculturel est triple. Premièrement, il permettra une mesure objective de la dimension comportementale de la compétence interculturelle, ce qui n'a guère été réalisé jusqu'ici. Deuxièmement, il répond à l'appel à plus de recherches expérimentales en management international (Zellmer-Bruhn, Caligiuri, & Thomas, 2016). Enfin, il développe une approche par le jeu en management interculturel, ce qui n'a encore que très rarement été fait, alors que les jeux sérieux attirent une attention toujours croissante en tant qu'outils innovants et performants pour l'apprentissage et la gestion de connaissances (Michel, Kreziak & Heraud, 2009 ; Vallat et al., 2016).

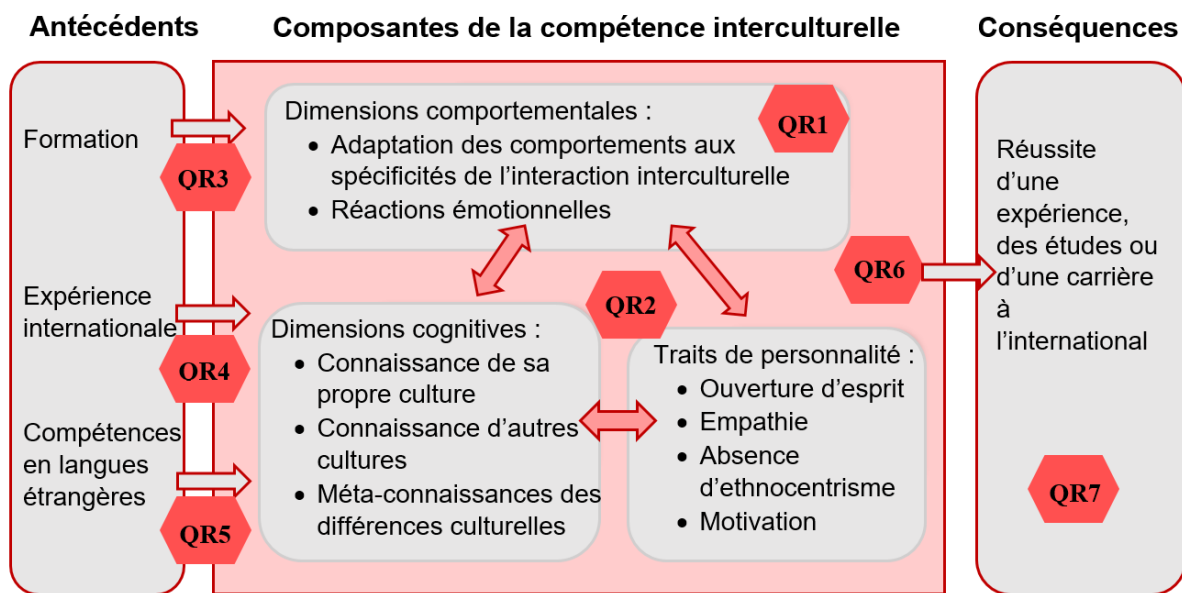
Le contexte imaginé pour *LINK the serious game*® est celui d'une équipe virtuelle de projet d'innovation. Les projets d'innovation sont de plus en plus souvent menés au sein d'équipes internationales, dispersées à travers le monde. Pour que la collaboration au sein de ces équipes puisse être fructueuse, la Compétence Interculturelle (CI) des membres de ces équipes est indispensable (Molinsky et al., 2012 ; Stahl et al., 2010). Le concept de CI a donné lieu à une très vaste littérature, en sciences de la communication, langues et civilisations, et en management international. Une grande diversité d'approches et de vocabulaire coexistent, mais une conceptualisation qui s'est progressivement imposée aborde la CI à travers ses composantes. Celles-ci peuvent être regroupées en trois catégories : les traits de personnalité vus comme liés à la CI (ex. l'ouverture d'esprit, l'empathie, l'absence d'ethnocentrisme, etc.), les connaissances des différences culturelles, et l'adaptation du comportement à celles-ci. On considère généralement que la formation, mais surtout l'expérience internationale contribuent à la CI. La figure 1 résume les antécédents, composantes et conséquences de la CI.

Les mesures existantes les plus courantes de la CI comportent toutes des biais importants, ce qui jette également des doutes sur le concept lui-même, qui est insuffisamment fondé empiriquement (Van de Vijver & Leung, 2009) :

- Très fréquemment, la compétence interculturelle est réduite à une seule de ces composantes : les traits de personnalité et attitudes liés à la compétence interculturelle, mesurés par des outils méthodologiques comme le MPQ (Multicultural Personality Scale, Van Oudenhoven & Van der Zee, 2002). Or, le lien entre cette composante et les autres n'a rien d'évident : dans notre recherche récente, seulement 17% des

connaissances interculturelles sont expliquées par ces traits de personnalité (Bartel-Radic & Giannelloni, 2017) ;

- Les mesures portant sur la notion d'intelligence culturelle (Ang et al., 2007 ; Thomas et al., 2008), définie de manière identique à la CI, sont certes multidimensionnelles (et portent donc aussi sur les dimensions cognitive et comportementale), mais le faible nombre de questions par dimension est autoévalué et non situé ;
- Dans la pratique davantage que dans la recherche, la compétence interculturelle est souvent considérée comme résultant « automatiquement » de l'expérience internationale. Or, nos recherches aboutissent à un lien à hauteur de 5% environ, donc extrêmement faible (Bartel-Radic, 2014).



Questions de recherche :

QR 1: Comment **mesurer les dimensions comportementales** de la compétence interculturelle de manière **expérimentale** ?

QR 2: Dans quelle mesure les **traits de personnalité** influencent la compétence interculturelle ?

QR 3: Quels types de **formations** impactent la compétence interculturelle ?

QR 4: Sous quelles conditions **l'expérience internationale** augmente-t-elle la compétence interculturelle ?

QR 5: Quand et dans quelle mesure les **compétences en langues étrangères** augmentent la compétence interculturelle ?

QR 6: La compétence interculturelle **diffère-t-elle selon les cultures** ?

QR 7: Dans quelle mesure la compétence interculturelle mène à la **réussite d'une expérience, des études ou d'une carrière à l'international** ?

Figure 1 : Antécédents, composantes et conséquences de la compétence interculturelle

Au niveau de la recherche, l'objectif du projet InterCCom est de contribuer à la réponse aux questions de recherche ci-dessus en développant un jeu sérieux informatisé. Ce jeu évite les biais méthodologiques des méthodes existantes et évalue connaissances et comportements interculturels en situation. Les résultats récoltés grâce à ce jeu permettront de répondre aux questions de recherche ci-dessus. Chacune des questions de recherche porte sur des gaps théoriques actuels dans le champ du management interculturel.

Concrètement, dans *LINK the serious game*®, le joueur prend le rôle du coordinateur d'une équipe fictive et virtuelle de projet d'innovation, équipe composée de personnes de diverses nationalités et dispersées autour du globe. Le joueur doit coordonner le travail de l'équipe, en adaptant son comportement à leurs cultures nationales respectives. En fonction des styles de communication et de management adoptés, les membres virtuels de l'équipe contribueront plus ou moins au travail de l'équipe, et à sa performance. Le scénario sera construit sur une succession de scénettes correspondant à la technique des « incidents critiques » (Flanagan, 1954), de brèves histoires relatant des situations interculturelles. Ces scénettes sont susceptibles d'être interprétées différemment selon les cultures car elles sont fondées sur les dimensions de la culture développées par Hofstede (1980, 2001), Hall et Hall (1990) et Hampden-Turner et Trompenaars (2003), relatives à la vision du groupe, de la hiérarchie, du temps, de la compétition, de la règle, etc. La technique des incidents critiques permet de capter des connaissances tacites liées à la CI (Johnson et al., 2006). L'utilisation de cinq incidents critiques dans des recherches préliminaires au projet (Bartel-Radic, 2014 ; Bartel-Radic & Giannelloni, 2017) a été vue comme une perspective prometteuse pour le champ. Dans ce projet, il s'agit d'aller au-delà et de les utiliser comme base pour une méthodologie quasi-expérimentale.

La méthode utilisée pour développer les scénettes et le scénario du jeu est fondée sur trois processus complémentaires :

- Des entretiens et des focus groups avec des responsables et membres d'équipes projets d'innovation internationaux ont été menés et analysés pour identifier des « incidents critiques » intéressants pour le jeu ;
- Lors des workshops collaboratifs de l'équipe projet interdisciplinaire, et très internationale, le théâtre d'improvisation a été utilisé comme méthodologie pour développer davantage de scènes, et pour échanger sur les réactions des porteurs de telle ou telle culture face à ces situations ;
- Pour les deux phases du travail collaboratif sur le jeu, c'est-à-dire également lors de la conception de la plateforme et du jeu, les méthodologies du design thinking (Dorst, 2011 ; Chanal & Merminod, 2018) ont été mobilisées. L'équipe projet a ainsi défini quatre *persona* de joueurs, travaillé sur les attentes et motivations respectives, avant de décrire le jeu sous l'angle de l'expérience utilisateur.

Le point commun entre les différentes applications de la collection *i-Team Games*® est que le joueur est un coordinateur et/ou membre d'une équipe virtuelle internationale. Spécificité du projet par rapport à d'autres qui existent dans le domaine, le joueur joue ainsi son propre rôle (en termes d'âge, de sexe, de nationalité, de lieu de travail...), ce qui contribue fortement au réalisme du jeu. Les échanges avec les autres membres (virtuels) de l'équipe simulée dans le jeu se font via les outils de communication digitaux, la messagerie électronique ou instantanée, le partage de documents, et la visioconférence (simulée à l'aide de brèves vidéos).

Si ces expérimentations permettront de capter les comportements des participants au sein d'équipes internationales virtuelles vis-à-vis de problématiques de recherche variées, on ne peut pas pour autant les qualifier d'expérimentations au sens strict du terme. Revenons sur les principes de l'expérimentation évoqués dans ce chapitre. Il est clair que nous nous trouvons

ici dans le cas d'une expérimentation clinique, où l'environnement du participant est reconstitué artificiellement, et ceci pour pouvoir au maximum maîtriser les différentes variables, et comparer un grand nombre d'individus. Deux autres principes clés sont « l'administration d'un traitement », une variable qu'on fait varier au cours de l'expérimentation, ainsi que l'existence d'un groupe de contrôle pour lequel cette variable de change pas. Le jeu *LINK the serious game*® comprend une évaluation de la dimension comportementale de la compétence interculturelle en début et fin de jeu, ainsi que des séquences de jeu suivis d'explications sous forme de cours au milieu. Une possibilité d'utilisation est donc de séparer l'échantillon en deux, avec une partie qui parcourt l'ensemble du jeu, l'autre qui ne réalise que les évaluations en introduction et en conclusion. Ce protocole permettrait de tester l'hypothèse qu'une formation à la collaboration interculturelle sous forme de jeu sérieux augmente la compétence interculturelle des participants. Mais les problématiques les plus intéressantes (cf. figure 1) sont bien plus difficiles à tester sous forme expérimentale. Les questions de recherche 1 et 2 portent sur la nature comportementale de la compétence interculturelle, mais n'impliquent pas de « traitement administré », ni même un test d'hypothèse systématique. Pour les questions de recherche 3 et 4 relatives aux facteurs augmentant la compétence interculturelle, deux possibilités de présentent : 1) utiliser la mesure de la compétence interculturelle par l'expérimentation à une seule reprise, et comparer des sous-groupes selon leur expérience internationale préalable, ou encore leur formation ; on est alors face à une pré-expérimentation tout comme à une quasi-expérimentation. Ou alors, 2) relever la compétence interculturelle grâce à l'expérimentation à deux reprises, avant ou après une expérience internationale ou une formation. Ce design de recherche est alors une (presque) vraie expérimentation, avec un groupe de contrôle (mais pas attribué de manière randomisé) et un « traitement » administré entre les deux mesures.

CONCLUSION

Quelle que soit la variable que le chercheur souhaite tester, ce dernier sera toujours pris entre le marteau et l'enclume, soit prisonnier de sa volonté de reproduire au plus près une réalité dans un environnement clinique, soit esclave de sa propension à vouloir déterminer l'environnement réel le plus propice à une expérimentation donnée. Il n'existe aucune « meilleure solution » en matière de recherche expérimentale. L'essentiel c'est de conserver une intégrité intellectuelle solide et d'accepter de faire une force d'une faiblesse, en d'autres termes de voir dans les limites d'une recherche une source de pistes d'explorations futures. L'environnement clinique a ceci de formidable qu'il permet d'être en contrôle de l'ensemble des variables endogènes et exogènes que l'on souhaite mobiliser pour mesurer ce que l'on souhaite mesurer. Et le terrain réel a ceci de merveilleux qu'il permet de tester ce que l'on veut tester en étant certain que toutes les incidences seront représentées. Au début du 18^e siècle déjà, Gottfried Wilhelm Leibniz faisait références aux « petites perceptions », toutes ces influences, suggestions, interférences et autres stimuli qui gouvernent nos perceptions comme nos choix sans pourtant que l'on soit en mesure de les détecter. Avec les progrès de la Science, nous savons désormais que l'homme est en mesure de capter 40 stimuli à la fois, issus de son environnement, et ce sur les milliers qui lui parviennent pourtant. Il ne verra donc jamais tout ce qui est exposé à ses yeux, ne sentira jamais tous les parfums qui lui parviennent, ne goûtera que ce qui est sucré, salé, amer ou acide, n'entendra que certains sons et ne touchera que ce qui sera à sa portée. L'essentiel lui échappera, comme il lui a toujours

échappé. D'où l'importance de la recherche expérimentale. Le chercheur doit donc passer par une école de frustration dans laquelle il devra avant tout déterminer ce qu'il ne pourra pas mesurer. Par la clinique, il réduira le nombre de ses variables mais atteindra des résultats d'une grande précision. Par le terrain, il augmentera le nombre de ses variables mais atteindra des résultats nécessairement biaisés. En Management International, le dilemme est peut-être encore plus marqué car à la complexité de l'humain, on rajoute celle de la culture. Tester des variables culturelles par la clinique permet de cerner des comportements stéréotypés. Mais ils peuvent être factices car surjoués. Tester des variables culturelles par le terrain permet de laisser libre cours aux réactions naturelles. Mais les personnalités individuelles prennent alors souvent le pas sur la culture, ce qui empêche de déterminer clairement ce qui est issu de quoi. La solution réside sans doute dans une approche hybride, une méthodologie mixte alliant la clinique et le terrain. Mais surtout, seuls le temps et la ténacité des chercheurs permettront de mieux comprendre l'Humanité dans ce qu'elle a de plus diversifié comme de plus identique. Et croire à cet avenir, c'est peut-être d'ores et déjà donner raison à Arthur Schopenhauer qui, en 1819, écrivait que « *ainsi la tâche n'est point de contempler ce que nul n'a contemplé mais de méditer comme personne n'a encore médité sur ce que tout le monde a devant les yeux.* »

BIBLIOGRAPHIE

- Ang, S., Van Dyne, L., Koh, C., Ng, K. Y., Templer, K. J., Tay, C., & Chandrasekar, N. A. (2007). Cultural Intelligence: Its Measurements and Effects on Cultural Judgment and Decision Making, Cultural Adaptation and Task Performance. *Management and Organization Review*, 3(3): 335-371.
- Bartel-Radic, A. (2014). La compétence interculturelle est-elle acquise grâce à l'expérience internationale ?" *Management International*, 18 (Special Issue): 194-211.
- Bartel-Radic, A., & Giannelloni, J.-L. (2017). A renewed perspective on the measurement of cross- cultural competence: An approach through personality traits and cross-cultural knowledge. *European Management Journal*, 35(5): 632-644.
- Binet, Denant-Boemont & Hammiche (2019), The Incidence of Price Information on Water Consumption Under Complex Tariff Scheme: A Laboratory Experiment, 6th French Association of Environmental and Resource Economists, University of Rennes 1.
- Caligiuri, P. M., & Phillips, J. (2003). An application of self-assessment realistic job previews to expatriate assignments. *International Journal of Human Resource Management*, 14(7): 1102-1116.
- Chanal, V., & Merminod, V. (2018). Comment adresser les problèmes pernicieux de manière créative avec le design thinking? *Management international* (à paraître).
- Cortes-Cornax, M., Front, A., Rieu, D., Verdier, C., & Forest, F. (2016). ADInnov: An Intentional Method to Instil Innovation in Socio-Technical Ecosystems. In International Conference on Advanced Information Systems Engineering CAiSE 2016: *Advanced Information Systems Engineering* (pp. 133-148). Springer, Cham.
- Davis, D. D., & Holt, C. A. (1993). *Experimental economics*. Princeton university press.
- Dorst, K. (2011). The core of 'design thinking'and its application. *Design studies*, 32(6), 521-532.
- Flanagan, J. (1954). The critical incident technique. *Psychological Bulletin*, 51(4): 327-358.
- Ferraro, P. J., & Price, M. K. (2013). Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment. *Review of Economics and Statistics*, 95(1), 64-73.

- Forest, F., Lavoisy, O., & Chanal, V. (2009). Integrated Scenario-based Design Methodology for Collaborative Technology Innovation, *The Future of Innovation (Int. Journal of Innovation Management)*, June, Vienn, 99.
- Grawitz, M. (1993). *Méthodes des sciences sociales*. Précis Dalloz, 9^e édition.
- Greenwood, E. (1972). Das Experiment in der Soziologie. In: König, R. (Ed.) *Beobachtung und Experiment in der Sozialforschung*. 8^e ed. Köln: Kiepenheuer& Witsch, 171-220.
- Hall, E. T., & Hall, M. R. (1990). *Guide du comportement dans les affaires internationales : Allemagne, Etats-Unis, France*. Paris, Seuil.
- Hampden-Turner, C., & Trompenaars, F. (2003). *Riding the waves of culture*. London, Nicholas Brealey.
- Harrison, G. W., List, J. A. (2004). Field experiments. *Journal of Economic literature*, 42(4) : 1009-1055.
- Hill, A. B. (1965). The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine*, 58(5), 295-300.
- Hofstede, G. (1980). *Culture's consequences: International differences in work related values*. London, Sage.
- Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations across Nations*. Thousand Oaks, CA, SAGE.
- Jacquemet N., L'Haridon, O., & Morin, P. (2013) Economie expérimentale et comportements. Introduction, *Revue française d'économie*, 28(2) : 3-8. DOI : 10.3917/rfe.132.0003
- Johnson, J. P., Lenartowicz, T., & Apud, S. (2006). Cross-cultural competence in international business: toward a definition and a model. *Journal of International Business Studies*, 37(4): 525-543.
- Igalens, J., & Roussillon Soyer, C. (2019) Les designs quasi-expérimentaux. In : Chevalier, F., Cloutier, L.M., & Mitev, N., *Les méthodes de recherche du DBA*, Paris : EMS, 251-271.
- Leibniz, G. W. (1703). Les petites perceptions. In : *Nouveaux essais sur l'entendement humain*, II, chapitre 1. Paris: Garnier (96-97).
- Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic perspectives*, 21(2), 153-174.
- Meyer, M. (2008). La méthode des scénarios : un outil d'analyse et d'expertise des formes de communication dans les organisations, *Études de communication*, 31 | 2008, 133-156.
- Michel, H., Kreziak, D., & Heraud, J.-M. (2009). Evaluation de la performance des Serious Games pour l'apprentissage : Analyse du transfert de comportement des élèves virtuels de Vacheland. *Systèmes d'information et management*, 14(4) : 71-86.
- Molinsky, A., Davenport, T. H., Iyer, B., & Davidson, C. N. (2012). Three Skills Every 21st Century Manager Needs. *Harvard Business Review*, 90(1-2), 139-143.
- Mouillot, Ph. (1999). *Fondements épistémologiques, théoriques et expérimentaux de la suggestion subliminale par le biais des images subliminales : Une approche de communication commerciale*. Thèse de doctorat, Université de Nice - Sophia-Antipolis.
- Mouillot, Ph., Drillon, D., & Montargot, N. (2018). Mosaïque des intelligences dans l'organisation : Éléments de distinction entre hauts potentiels, talents et HQI. *Revue Internationale de Psychosociologie et de Gestion des Comportements Organisationnels*, XXIV(58), 171-196.
- Pornpitakpan, C. (1999). The effects of cultural adaptation on business relationships: Americans selling to Japanese and Thais. *Journal of International Business Studies*, 30(2): 317-337.

- Roth, A., Prasnikar, V., Okuno-Fujiwara, M., & Zamir, S. (1991). Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study. *American Economic Review*, 81(5), 1068-1095.
- Schnell, R., Hill, P. B., & Esser, E. (2018). *Methoden der empirischen Sozialforschung*. Oldenbourg : De Gruyter.
- Schopenhauer, A. (1819). *Le monde comme volonté et comme représentation*. Traduction A. Burdeau, PUF, 1992, p. 58.
- Spearman, C. (1904). "General Intelligence", Objectively Determined and Measured. *American Journal of Psychology*, 15(1): 72-101.
- Stahl, G. K., Maznevski, M. L., Voigt, A., & Jonsen, K. (2010). Unraveling the effects of cultural diversity in teams: A meta-analysis of research on multicultural work groups. *Journal of International Business Studies*, 41(4): 690-709.
- Thomas, D. C., Elron, E., Stahl, G., Ekelund, B., Ravlin, E., Cerdin, J.-L., Poelmans, S., Brislin, R., Pekerti, A., Aycan, Z., Maznevski, M., Au, K., & Lazarova, M. (2008). Cultural Intelligence: Domain and Assessment. *International Journal of Cross-Cultural Management* 8(2): 123-143.
- Vallat, D., Bayart, C., & Bertezene, S. (2016). Serious games in favour of knowledge management and double-loop learning? *Knowledge Management Research & Practice*, 14(4): 470-477.
- Vijver, F. J. R., & Leung, K. (2009). Methodological Issues in Researching Intercultural Competence. In Deardorff, D. K. (Ed.), *The SAGE Handbook of intercultural competence*. (404- 418), Thousand Oaks, SAGE.
- Van Oudenhoven, J. P., & Van der Zee, K. I. (2002). Predicting multicultural effectiveness of international students: The Multicultural Personality Questionnaire. *International Journal of Intercultural Relations* 26(6): 679-694.
- Van Witteloostuijn, A. (2015). Toward experimental international business: Unraveling fundamental causal linkages. *Cross Cultural Management*, 22(4), 530-544.
- Zellmer-Bruhn, M., Caligiuri, P., & Thomas, D. (2016). From the Editors: Experimental designs in international business research. *Journal of International Business Studies* 47(4): 399-407.

AUTEURS

Anne Bartel-Radic

Professeur des universités en management international
Univ. Grenoble Alpes, Grenoble INP, CERAG, F-38000 Grenoble, France

Philippe Mouillot

Maître de Conférences HDR en Sciences de gestion
IAE de Poitiers, France

Danielle Taylor

Doctorante en management international
Univ. Grenoble Alpes, CERAG, F-38000 Grenoble, France

Experimental Methods in International Management Research

ABSTRACT

Experimental methods have long been used in the natural sciences to test a hypothesis by controlling experimental conditions. These methods have also been used in economics and management in recent decades, and have become a reference in some sub-disciplines such as marketing. However, experiments are rare in international management research. In this chapter, we present the difficulties, but also the advantages of the experimental methodologies in this sub-discipline through the example of the InterCCom project, which deals with intercultural competence. Thanks to the development of a computerized serious game for this project, the behaviors of the members of a virtual, international team are measured. Various research questions are outlined and answered.

KEYWORDS: Experimental methods, experimentation, international management, intercultural competence

INTRODUCTION

The purpose of this chapter is to provide researchers and social science students with an introduction to experimental methodologies. These methodologies originally come from the natural sciences such as physics, chemistry or medicine. In recent decades, they have been widely used in economics and management science, particularly in the sub-discipline of marketing. On the other hand, in international management, contributions remain relatively rare, whereas the best journals of the discipline have called for a significant increase of this research.

This chapter is structured in four parts:

1. Clinical experimentation as used in medicine or neuroscience.
2. The establishment and use of experimental methods in economics and management science.
3. Application to the international management field and the inherent difficulties in this sub-discipline.
4. The presentation of a research project in intercultural management, and how it mobilizes experimental methods.

1. EXPERIMENTAL METHODS: SOME FUNDAMENTALS

In a broad sense, any empirical data collection that aims to test a hypothesis, whatever the mode of data collection (observation, interviews, focus group, or questionnaire), can be described as experimentation. The hypotheses to be tested include links between one or more causes, and a consequence. A link between two phenomena can be qualified as causal if the following criteria are met (Hill, 1965):

- the correlation must be strong;
- this correlation must be shown in different contexts, in studies conducted on different populations;
- the cause should have specific consequences;
- the cause should precede the consequence in time;
- it must be possible to find a plausible explanation of how the cause influenced the consequence;
- and, in general, it must be possible to find a monotonic function between the cause and the consequence (*more* causes should result in *more* consequences).

There are various experimental methods including “real” experimental designs, quasi-experimentation and other data collection techniques. According to Greenwood (1972: 178), true experimental designs must meet the following criteria:

- they aim to test a hypothesis;
- this test goes through a precise and controlled research protocol;
- and the experimental conditions are controlled using variable elimination techniques, constant conditions and random sampling.

Only research designs that include two measures, such as before and after an event or treatment proposed to the test group (e.g. inclusion of a variable), and that have a control group that has not undergone the same treatment abide by these conditions. Thus, in pharmacology, for example, a drug meant to cure a certain pathology can be tested through a process where the test group of patients is given the drug while the control group is given a placebo. Both groups are given an evaluation of their state of health before and after the treatment.

Thus, we distinguish the experimental protocols that take place in someone’s normal environment (“field experiment”) from experimental protocols that take place in a dedicated place such as a laboratory (i.e. “clinical experimentation”). Studies taking place in a person’s normal environment have a higher external validity because people’s behaviors will be less biased. But not only are these field trips costlier in terms of time and resources, they also do not allow controlling contextual variables with the same precision. Clinical experiments are thus usually given greater internal validity, even if they also have limitations (Schnell, Hill & Esser, 2018).

The very foundation of clinical experimentation rests on three essential pillars:

- An artificial reproduction that is as faithful as possible to the environment that the experiment is testing.
- A judicious choice of tools and methods to mobilize to achieve one's goals.
- A research question that is clear, measurable and relevant.

For the first pillar to be stable, a noted technique in experimentation is that reality should first be observed at length. Indeed, in terms of intercultural management, immersion in the country's culture, frequent sharing with its representatives and specific work on stereotypes can prove to be very effective. For the second pillar to be stable, it is sufficient to justify each decision; to do so, we should explain the reasoning for each choice by constantly employing the word "because." As for the third pillar, only one question merits recurrence: what do we want to measure?

Quasi-experimental designs are qualified as research designs in which an experimental group and a control group are compared, but where the classification of an individual in one group or the other is not randomly distributed. Generally, the observed individuals choose the group they want to belong to, which creates biases. Within these quasi-experimental designs, we distinguish "natural experiments" from "quasi-experiments." In the former, manipulation of the variable occurs without action by the researcher (such as the decision to reduce speeds from 90 km/h to 80 km/h on national roads in France in 2018), while in the latter, the researcher deliberately manipulates the variable (by introducing a new public policy in only one or a few geographical areas that are expressly chosen to verify its impact, for example).

The use of experimental methods in the social sciences faces a number of limitations. We need very large samples (for example 1500 people if we want to be able to generalize the results to the population of a country like Germany, Schnell, Hill & Esser, 2018: 203). Samples should be comprised of people ideally chosen at random, which requires significant resources. Furthermore, and most importantly, many variables studied in the social sciences, such as culture, intelligence, or socio-professional categories cannot easily be manipulated during the experiment, and especially do not allow people to be randomly assigned to one of the two groups.

Take the case of a measure of intelligence. The general factor (or g factor - Spearman, 1904) consists of agreeing on a conceptualization of the object of research, both in its definition and in its recurrent structure, in order to be able to quickly identify any dissonance. Here, it is essential to distinguish the subject capable of solving a problem because s/he has already solved it, because s/he has observed its solution, or because s/he is using his/her unique intelligence. Thus, we distinguish memory, learning, and "pure" intelligence (Mouillot, Drillon, & Montargot, 2018). In a study on intelligence, we could lament that the only factor tests still designed today are matrices. But the matrices do not eliminate the cultural aspect; they implement many cognitive dimensions (which is their strength), among which are working memory and verbalization, for example. There is no doubt that such an approach leads to differential psychology. This makes the object of study the existence of stable characteristics in individuals, which distinguish one individual from another.

Indeed, creating a clinical experiment not only makes it possible to reveal differences (even concerning less apparent dimensions than hair color, for example), but also to measure this stability. The stability of these differences is the basis for the prognosis. Before any creation of a clinical experimentation tool, it is therefore essential to agree on the essence and the boundaries of what is being measured. If, in the context of intelligence for example, researchers define intelligence as the ability to solve a problem that has never before been encountered, then they could vary the clinical elements of the test according to degrees of difficulty or the g factor of the novelty of the problem to be solved being systematically clinically present.

Let's talk about the test itself. This technique is generally used to establish numerical differentiations, thanks to statistical tools such as dispersion, standard deviation, etc. However, it is also possible to use testing methods to place the subject in an atypical experimental situation, which is nevertheless effective. Take the case of the Thematic Aperception Test (T.A.T.) that Henry Murray created in 1935. In clinical psychology, the T.A.T., frequently used in parallel with Herman Rorschach's Test, consists of a series of storyboards showing scenes from everyday life. Unlike the Rorschach Test where the patient is questioned in order to know what s/he "sees" in the inkblots presented to him/her, in the T.A.T. the patient is asked to tell a story about what s/he sees. S/he therefore uses a multitude of complex and subconscious elements that guide him/her in his daily life, but which, in this case, are mobilized artificially through a clinical suggestion. It then becomes very interesting to use variants of this type of test in order to know the personal and private feelings of a subject. So, for example, rather than ask him/her if s/he thinks that a Mercedes is a younger or more mature brand, or if milk makes people fat, presenting the image of two different silhouettes and asking which drinks milk and which drives a Mercedes is enough. Subconscious associations will do the rest.

Finally, from the results of one test, factor analysis allows researchers to isolate factors that are common to several tests, which in turn allows the interpretation of numerical or qualitative findings. In this case, we are really at the heart of comparative clinical analysis with the control group in opposition to the experimental groups. Factor analysis makes it possible to clearly identify the variants and their impact on the subjects belonging to the control group. For example, if one wishes to set up a protocol that makes it possible to artificially and subconsciously influence subjects to provide certain answers to a questionnaire, it is essential to start by submitting the questionnaire to a control group to be able to then indicate the variations between the groups.

2. EXPERIMENTATION IN ECONOMICS AND BUSINESS ADMINISTRATION

In 1993, Davis and Holt wrote: "A small but increasing number of economists have begun to make use of experiments under economic conditions." The situation has dramatically changed since then, and behavioral economics has become a vast and widely recognized field of research. Experiments are used to test economic theories, such as financial decision-making, public goods provision, market equilibrium, public policy and other variations in the environmental economics (Jacquemot, l'Haridon & Morin, 2013; Lunn & Choidealbha, 2018; Ferraro and Price, 2013; Binet, Denant-Boemont & Hammiche, 2019). In economics,

experimentation has attained such success because lab and field experiments allow researchers to observe and analyze the specific determinants of economic behavior (Jacquemot, l'Haridon & Morin, 2013). Harrison and List (2014) further specify that a lab experiment is an efficient method in order to create counterfactual scenarios. Lunn and Choidealbha (2018) go so far to state that a well-designed lab study can be the best method for answering policymakers' research questions.

In management research, the use of experimental methods has also been reinforced (see Igalens & Roussillon Soyer, 2019), but later, and to a lesser extent. Specifically, marketing and behavioral finance are the fields of research where they are the most frequently mobilized. In both, research questions may reflect consumer or purchasing choices (for example, stocks).

Research designs that only provide measurement at a time t , and therefore no measurement before or after the event, are qualified as pre-experiments. A method of collecting empirical data that is very common in management science is the questionnaire. In these "ex-post-facto-designs," the researcher can certainly study correlations. However, s/he cannot manipulate the independent variable, which does not allow the elimination of a certain number of alternative explanations as in true experimentation.

Contrarily, clinical experimentation consists of reproducing a real situation in a setting or artificial environment while maintaining the maximum fidelity for the reality from which it comes in order to test a variable or to measure a variation (*delta*). For this, the researcher will either start from real statistical data collected from longitudinal observations, or s/he will observe the behavior of a group called the "control" in order to have a comparative basis before observing so-called "experimental" groups, thus with modifications of the measured variable. For example, based on consumption statistics, the researcher can recreate an environment close to that of a supermarket and provide the subjects with a shopping list before observing the way they walk around, their hesitations, etc. Another example, if you want to measure the coherence of a website's architecture, the researcher could simply place topics on a screen and ask the subjects to find some certain information. The researcher can then observe the path the subjects take, this time virtually.

Clinical experimentation can only help study the usual laws of behavior by systematically varying the factors that influence it. However, some of these factors can only be changed by using the differences between individuals and by comparing the answers given by a certain group of subjects to the answers given by another correctly chosen group. The latter must be of equivalent quantitative and qualitative composition. It is therefore essential to identify the important factors that are at the origin of the studied behavior in order to either modify them, withdraw one or more of them, or even to add one or more of them. This is why the use of control and experimental groups, which must be similar in their representations and content, is so important. Without systemic, stable and recurrent reference, it is impossible to observe a salience, a digression or an anomaly.

The following example is taken from the doctoral research of one of the authors who specialized in neuromarketing (Mouillot, 1999). The aim was to measure the degree of suggestibility of two experimental groups that received a subliminal suggestion, in other words a subconscious influence. Why two groups? Because it was necessary to check the results of

the first one with a second experiment. Indeed, if the control group serves as a reference before the clinical variable is mobilized, the results from an experimental group must be confirmed by a second experiment organized under the same circumstances but with different subjects in order to highlight any result variation. If both experimental groups produce data, whose proximity is considered acceptable - this interval must be defined upstream of the experiment - then the influence of the clinical variable becomes comparable to the control group. On the other hand, if the results provided by the two experimental groups are frankly divergent or convergent but judged to be insignificant, then the use of a third experimental group is necessary to identify the element(s) that caused this divergence.

In the present example, the research hypothesis was that natural behavior could be subconsciously influenced, thus modified, for at least 30% of the subjects participating in the experimentation, with more or less 5 points of variation. For this purpose, we used a control group consisting of 159 subjects and two experimental groups consisting respectively of 95 and 58 subjects. Methodologically, the number of control subjects was sufficiently close to the number of experimental subjects (159 vs. 153) to allow comparison. A questionnaire was submitted to the control group (159 subjects) in order to know the degree of market penetration of non-alcoholic drink brands without subconscious suggestion and independently of any sensory suggestion. The subjects completed the questionnaire in a room where there was no reference to any brand. Unsurprisingly, for this study, which took place in France and among French consumers, Coca-Cola, Evian, Orangina, Perrier and Pepsi-Cola collected the highest number of answers.

Then, the first experimental group was mobilized (95 subjects). This group was presented three video clips aimed at subliminally suggesting Teisseire, a French flavored syrup brand. The brand image was inserted and projected at a speed of 1/25th of a second. The group was invited to watch these clips in order to choose which one would be most relevant to accompany an advertising campaign aimed at fighting AIDS. While both the first experimental group and the control group named most of the same brands, Teisseire reached 4th position and Pepsi-Cola was taken out of the rankings. With 33.68% of the answering behaviors having been modified, our hypothesis was validated. It only remained to verify with a second experimental group that these results were methodologically acceptable. Seven months later, a second experimental group was formed (58 subjects) and submitted to the same protocol. After viewing the clips, Teisseire reached 5th position in the questionnaire answers, this time replacing Perrier. In this second case, 27.33% of the responses were modified. We hypothesized an artificial behavioral change within a range of 25 to 35%, and our two results were within this range. Since Teisseire was never mentioned by the control group subjects, and a Chi-square test had validated our statistics, this result encouraged us to think that artificial subconscious suggestion had a real impact on the randomly-chosen 18 to 25-year-old subjects, whatever their gender.

3. EXPERIMENTATION IN INTERNATIONAL MANAGEMENT

Certain fields such as economics, psychology and marketing increasingly use experimental research designs, to the point where they have become almost mandatory for high-level publications. By contrast, in international management, Zellmer-Bruhn, Caligiuri, and Thomas (2016) found that less than 1% of more than 900 empirical studies published in the *Journal of International Business Studies* (JIBS) use experimental methods. This is mainly due, among other things, to both the difficulty of designing appropriate research protocols that can be randomly assigned to participants and to the challenges related to sampling.

By 2016, experimentation was reportedly only present in a limited number of business and management subjects, including marketing/ advertising consumer behavior, sales communication, venture capitalists' (VCs') decision-making, cultural differences in decision-making, and empowerment and job satisfaction (Zellmer-Bruhn, Caligiuri & Thomas, 2016). However, it must be noted that few of these experiments were true experiments due to their non-randomized sampling approaches.

In the international business (IB) context, particular challenges to conducting experiments lie in the difficulties related to randomization and the nature of the samples. In other words, it is impossible for researchers to completely randomize their samples. As Zellmer-Bruhn, Caligiuri and Thomas (2016) convey, researchers cannot simply randomly assign countries to political economies, companies to globalization strategies, or country of origin to individuals, for example. Furthermore, student samples have been long viewed as inferior in the IB community, whereas they are often used in other disciplines such as psychology. While student samples may be appropriate for certain studies in IB, undergraduates cannot demonstrate proficient background and experience to respond sensibly to all experiments (Van Witteloostuijn, 2015). Furthermore, on the subject of samples, IB requires specified and varying cultural and institutional backgrounds, often with people from different geographic locations (Zellmer-Bruhn, Caligiuri & Thomas, 2016). This can represent a challenge simply in recruiting the appropriate participants.

In IB research, the most common experiments relate to individual or team-level outcomes. In international marketing, where experiments have long taken place, the study by Pornpitakpan (1999) provides an interesting example. Pornpitakpan (1999) concludes that Americans who adapt to two cultures and languages, in this case Thai and Japanese, have more positive sales outcomes even within their national contexts. In international economics, Roth, Prasnikar, Okuno-Fujiwara & Zamir (1991) employ an experimental design to compare bargaining behaviors among different cultures. They demonstrate that what an "acceptable" offer is depends on cross-country differences. In management, Caligiuri and Phillips (2003) conducted an experiment, randomly assigning expatriate participants to use or not a self-assessment decision-making tool. They showed that participants who received the tool, along with a realistic job preview (RJP), reported increased confidence in their decision to accept an international assignment and had greater success during said assignment.

In *intercultural* management (that is if one wants to go beyond the comparison of different cultures), clinical experimentation is traditionally interactional. Indeed, since interculturality necessarily implies the existence of a relationship between at least two people with different

personal, social and societal repositories, this approach aims to demonstrate the existence of interactions between two subjects. Any response is then considered to be a reaction to one or more stimuli, knowing that the behavior of a person cannot be changed without concomitant changes in his entourage.

However, in terms of clinical experimentation, it is all about interpretation, especially when one is working within the broad field of intercultural analysis. Without interpretation, any measure is meaningless, since it is neither put into perspective nor placed in a context, which in this case is experimental. Thus, when one speaks of the test method, without analysis of the standard deviation, an average could appear indicative of a tendency whereas it is not so. It is the same with dispersion. From a single cloud of points on which a large concentration would appear separated from a few isolated points, a study of consumption behaviors would be tempted to focus on the cloud of points whereas a psychological or intercultural study would be inspired to observe isolated points, which could potentially be respectively representative of psychological deviations or cultural specificities. In research, the very nature of the information that needs to be gathered to achieve the objective therefore commands the means employed to do so. Indeed, since 1993 when her book *Methods of Social Sciences* was published, Madeleine Grawitz has invited us to think about the fact that if it is possible to consider catching fish with a butterfly net, we certainly do not catch butterflies with hooks. Moreover, Friedrich Nietzsche also always maintained that the most precious riches were methods, this being a viewpoint that Rene Descartes could not deny either, for it is not the object that makes science but the method.

Despite the small number of studies mobilizing experimental methodologies in international management, multiple authors highlight its bright future. Zellmer-Bruhn, Caligiuri and Thomas (2016) conclude that experimentas are underrepresented in JIBS, but that they provide a clear opportunity to develop the evidence for causal relationships in international business research. Van Witteloostuijn (2015) encourages the development of an “experimental IB tradition” through the creation of web-based tools to further research, but also to increase the effectiveness and fun of IB teaching. In this way, we will next discuss such a tool created by the InterCCom project, that studies as well as teaches about intercultural teamwork.

4. THE INTERCCOM PROJECT: A PLATFORM FOR INTERNATIONAL MANAGEMENT EXPERIMENTATION

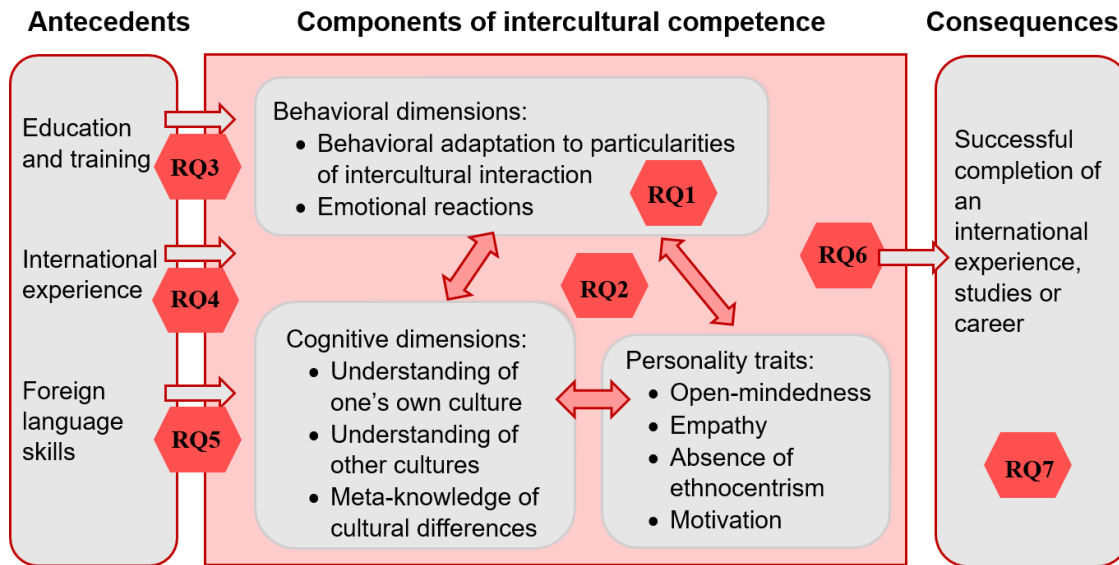
In this last part of the chapter, we present a project that has been conducted since January 2019 at Sciences Po Grenoble and University Grenoble Alps, which aims at building a serious inter-cultural management game called LINK the serious game®. The game be used as a quasi-experimentation tool in research because the choices made by the players are recorded. The game is implemented by a digital platform also developed within the project, the GenaGame® platform. On this platform, content (documents, images and videos) can be uploaded, and an interactive journey of the player through a content, depending on a scenario and underlying theoretical dimensions, can be defined. Other games in international management are thus under development. All of these games are grouped under the i-Team Games® brand.

The relevance of the InterCCom project within the field of intercultural management is threefold. First, it will allow an objective measure of the behavioral dimension of intercultural competence, which has not yet been achieved. Second, it responds to the call for more experimental research in international management (Zellmer-Bruhn, Caligiuri, & Thomas, 2016). Finally, it helps develop a game-based approach to intercultural management, which has only rarely been done, whereas serious games are attracting ever-increasing attention as innovative and effective tools for learning and management knowledge (Michel, Kreziak & Heraud, 2009, Vallat et al., 2016).

The context imagined for LINK the serious game® is that of a virtual team working on an innovation project. Innovation projects are more and more often carried out within international teams that are scattered throughout the world. For collaboration within these teams to be successful, the Intercultural Competence (IC) of those team members is essential (Molinsky et al., 2012, Stahl et al., 2010). The concept of IC has given rise to a vast literature in communication sciences, languages and civilizations, and of course international management. A great diversity of approaches and vocabulary coexist, but a conceptualization that has gradually imposed itself addresses the IC through its components. These can be grouped into three categories: personality traits seen as IC-related (e.g. open-mindedness, empathy, lack of ethnocentrism, etc.), knowledge of cultural differences, and the behavioral adaptation to the latter. Training, but especially international experience, is generally considered to contribute to IC. Figure 1 summarizes the background, components and consequences of IC.

The most common existing measures in IC all have important biases. This casts doubt on the concept itself, which is insufficiently empirically based (Van de Vijver & Leung, 2009):

- Very often, intercultural competence is reduced to only one of its components: personality traits and attitudes related to intercultural competence, measured by methodological tools such as the MPQ (Multicultural Personality Scale, Van Oudenhoven & Van der Zee, 2002). However, the link between this component and the others is not obvious. In fact, in recent research, only 17% of intercultural knowledge is explained by these personality traits (Bartel-Radic & Giannelloni, 2017).
- Measures on the notion of cultural intelligence (Ang et al., 2007, Thomas et al., 2008), defined in the same manner as IC, are certainly multidimensional (and therefore also include cognitive and behavioral dimensions). However, the small number of questions per dimension is self-assessed and not located.
- In practice more than in research, intercultural competence is often seen as “automatically” resulting from international experience. However, our research results in a demonstrated link of up to about 5%, which is extremely weak (Bartel-Radic, 2014).



Research Questions:

- RQ 1: How should **behavioral dimensions** of intercultural competence be measured **experimentally**?
- RQ 2: To what extent do **personality traits** influence intercultural competence?
- RQ 3: What types of **education and training** impact intercultural competence?
- RQ 4: Under what conditions does **international experience** increase intercultural competence?
- RQ 5: When and to what extent do **foreign language skills** increase intercultural competence?
- RQ 6: Does intercultural competence **differ among different cultures**?
- RQ 7: To what extent does intercultural competence lead to success in a person's **international or intercultural experience, studies or career**?

Figure 1: Antecedents, components and consequences of intercultural competence

At the research level, the objective of the InterCCom project is to answer the above research questions through the development of a serious computer game. This game avoids the methodological bias of existing methods and evaluates knowledge and intercultural behavior in various realistic situations. The results collected through this game and subsequently analyzed will answer the aforementioned research questions. Each of the research questions addresses current theoretical gaps in the field of intercultural management.

Specifically, in LINK the serious game®, the player assumes the role of the coordinator of both a fictional and virtual team of an innovation project, a team composed of people of various nationalities and scattered around the globe. The player must coordinate the work of the team, adapting their behavior to their respective national cultures. Depending on the communication and the management styles adopted, the virtual members of the team will more or less contribute to the team's work and its performance. The scenario will be built through a succession of scenes corresponding to the technique of “critical incidents” (Flanagan, 1954) or brief stories recounting intercultural situations. These scenes are likely to be differently interpreted according to different cultures because they are based on the cultural dimensions developed by Hofstede (1980, 2001), Hall and Hall (1990) and Hampden-Turner & Trompenaars (2003), which relate to the vision of the group, hierarchy, time, competition, rules, etc. The critical incident technique captures tacit knowledge related to IC (Johnson et al., 2006). The use of five critical incidents in the preliminary project research (Bartel-Radic, 2014, Bartel-Radic & Giannelloni, 2017) has been viewed as a promising prospect for the field. In

this project, the goal is to go beyond and use them as a basis for a quasi-experimental methodology.

The method used to develop the storyboards and the game scenario is based on three complementary processes:

- Interviews and focus groups with leaders and members of international innovation project teams were conducted and analyzed to identify “critical incidents” of interest to the game.
- During the collaborative workshops which brought together an interdisciplinary and very international project team, improvisation theater was used as a methodology to develop more scenes, and to discuss the reactions of the bearers of this or that culture to these situations.
- For the two phases of collaborative work on the game, that is to say also during the platform and game design, we mobilized design thinking methodologies (Dorst, 2011, Chanal & Merminod, 2018). The project team defined four personas of players and worked on the respective expectations and motivations before describing the game from the perspective of the users’ experiences.

The commonality between the different applications of the i-Team Games® collection is that the player is a coordinator and/or member of an international virtual team. A specificity of the project compared to others that exist in the field is such that the player plays his/her own role (in terms of age, sex, nationality, place of work, etc.), which greatly contributes to anchoring the realism of the game. The virtual exchanges with the other members of the team simulated in the game are done via digital communication tools, electronic or instant messaging, documents sharing and videoconference (simulated with the help of short videos).

If these experiments can capture the behavior of participants in virtual international teams vis-à-vis various research issues, we cannot qualify them as experiments in the strict sense of the term. Let's go back to the principles of experimentation mentioned in this chapter. It is clear that this is the case of a clinical experiment, where the environment of the participant is artificially reconstituted, in order to be able to control the various variables as much as possible and to compare a large number of individuals. Two other key principles are “the administration of a treatment or experiment,” a variable that is varied during the experiment, as well as the existence of a control group for which this variable does not change. The game LINK the serious game® includes an assessment of the behavioral dimension of intercultural competence at the beginning and the end of the game, as well as game sequences followed by explanations in the form of included lessons. A possibility of use is therefore to separate the sample in two, with one part going through the whole game, and the other one which only carries out the evaluations in both the introduction and the conclusion. This protocol would allow testing the hypothesis that training in intercultural collaboration in the form of serious play increases the participants’ intercultural competence. But the most interesting issues (see Figure 1) are much more difficult to test in an experimental form. Research questions 1 and 2 focus on the behavioral nature of intercultural competence, but do not imply “administered processing” or even a systematic hypothesis test. For research questions 3 and 4 on factors increasing intercultural competence, there are two possibilities: 1) to use the measure of intercultural competence through one-time experimentation, and to compare subgroups

according to their prior international experience or their training. We are then in the case of pre-experimentation as well as a quasi-experiment. Or, 2) raise intercultural competence through experimentation twice, before and after international experience or training. This research design is then an (almost) real experiment, with a control group (but not assigned in a randomized manner) and “processing” that is administered between the two measures.

CONCLUSION

Whatever the variable that the researcher wishes to test, the latter will always be stuck between the devil and the deep blue sea, either as a prisoner of his will to reproduce a reality in a clinical environment, or as a slave to his predisposition to determine what real environment would be the most conducive to welcome a given experiment. There is no “best solution” in experimental research. The key is to maintain a strong intellectual integrity and to accept the need to find strength in weakness, in other words to see a source of future exploration within the limits of a research study. A clinical environment is remarkable because it allows the researcher to be in control of all endogenous and exogenous variables that s/he wants to mobilize to measure what s/he ultimately aims to measure or understand. And the field is outstanding because it allows the researcher to test what s/he wants to test with the confidence that all effects will be represented. As early as the beginning of the 18th century, Gottfried Wilhelm Leibniz was referring to “small perceptions,” or all of those influences, suggestions, interferences, and other stimuli that govern our perceptions and choices without us being able to detect them. With the progress of science, we now know that a human being is able to capture 40 stimuli at a time from his/her environment among the thousands that yet reach him/her. S/he will never see all that is exposed to his/her eyes, s/he will never feel all the perfumes and odors that reach him/her, s/he will only taste what is sweet, salty, bitter or sour, s/he will only hear certain sounds, and s/he will only touch what will be within his reach. The essential will be out of reach, as has always been the case. This is the reason why experimental research is so important. The researcher must therefore go through a phase of frustration in which s/he must first of all determine what s/he *cannot* measure. The number of variables will be reduced but the researcher will reach results with a greater amount of precision. In reality, s/he will increase the number of variables but necessarily reach biased results. In international management, the dilemma is perhaps even more tangible because to the complexity of the human is added to that of culture. Testing cultural variables through a clinical approach can help identify stereotyped behaviors. However, those behaviors can be fake as they may be overplayed. Testing cultural variables in a real environment creates a space in which natural reactions can have free rein. On the other hand, individual personalities often take precedence over culture, which makes it difficult to clearly determine what comes from what. The ultimate solution probably lies in a hybrid approach, a mixed methodology combining clinical and field approaches. But above all, only time and the tenacity of researchers will make it possible to better understand Humanity, in its diversification but likewise in its similarities. To believe in this future, it is perhaps already right to rule in favor of Arthur Schopenhauer who, in 1819, wrote, *“thus the task is not to contemplate what no one has contemplated but to meditate like no one has ever meditated on what everyone has in front of them.”*

REFERENCES

- Ang, S., Van Dyne, L., Koh, C., Ng, K. Y., Templer, K. J., Tay, C., & Chandrasekar, N. A. (2007). Cultural Intelligence: Its Measurements and Effects on Cultural Judgment and Decision Making, Cultural Adaptation and Task Performance. *Management and Organization Review*, 3(3), 335-371.
- Bartel-Radic, A. (2014). La compétence interculturelle est-elle acquise grâce à l'expérience internationale ?" *Management International*, 18 (Special Issue), 194-211.
- Bartel-Radic, A., & Giannelloni, J.-L. (2017). A renewed perspective on the measurement of cross-cultural competence: An approach through personality traits and cross-cultural knowledge. *European Management Journal*, 35(5), 632-644.
- Binet, Denant-Boemont & Hammiche (2019), The Incidence of Price Information on Water Consumption Under Complex Tariff Scheme: a Laboratory Experiment, 6th French Association of Environmental and Resource Economists, University of Rennes 1.
- Caligiuri, P. M., & Phillips, J. (2003). An application of self-assessment realistic job previews to expatriate assignments. *International Journal of Human Resource Management*, 14(7), 1102-1116.
- Chanal, V., & Merminod, V. (2018). Comment adresser les problèmes pernicieux de manière créative avec le design thinking? *Management internationale* (in press/online first).
- Cortes-Cornax, M., Front, A., Rieu, D., Verdier, C., & Forest, F. (2016). ADInnov: An Intentional Method to Instil Innovation in Socio-Technical Ecosystems. In International Conference on Advanced Information Systems Engineering CAiSE 2016: *Advanced Information Systems Engineering* (pp. 133-148). Springer, Cham.
- Davis, D. D., & Holt, C. A. (1993). *Experimental economics*. Princeton university press.
- Dorst, K. (2011). The core of 'design thinking' and its application. *Design studies*, 32(6), 521-532.
- Ferraro, P. J., & Price, M. K. (2013). Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment. *Review of Economics and Statistics*, 95(1), 64-73.
- Flanagan, J. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), 327-358.
- Forest, F., Lavoisy, O., & Chanal, V. (2009). Integrated Scenario-based Design Methodology for Collaborative Technology Innovation, *The Future of Innovation (Int. Journal of Innovation Management)*, June, Vienna, 99.
- Grawitz, M. (1993). *Méthodes des sciences sociales*. Précis Dalloz, 9^e édition.
- Greenwood, E. (1972). Das Experiment in der Soziologie. In: König, R. (Ed.) *Beobachtung und Experiment in der Sozialforschung*. 8^e ed. Köln: Kiepenheuer & Witsch, 171-220.
- Hall, E. T., & Hall, M. R. (1990). *Guide du comportement dans les affaires internationales : Allemagne, Etats-Unis, France*. Paris, Seuil.
- Hampden-Turner, C., & Trompenaars, F. (2003). *Riding the waves of culture*. London, Nicholas Brealey.
- Harrison, G. W., List, J. A. (2004). Field experiments. *Journal of Economic literature*, 42(4), 1009-1055.
- Hill, A. B. (1965). The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine*, 58(5), 295-300.
- Hofstede, G. (1980). *Culture's consequences: International differences in work related values*. London, Sage.

- Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations across Nations*. Thousand Oaks, CA, SAGE.
- Igalens, J., & Roussillon Soyer, C. (2019) Quasi experimental design. In : Chevalier, F., Cloutier, L.M., & Mitev, N., *Research methods for the DBA*, Paris : EMS, 245-260.
- Jacquemet N., L'Haridon, O., & Morin, P. (2013) Economie expérimentale et comportements. Introduction, *Revue française d'économie*, 28(2) : 3-8. DOI : 10.3917/rfe.132.0003
- Johnson, J. P., Lenartowicz, T., & Apud, S. (2006). Cross-cultural competence in international business: toward a definition and a model. *Journal of International Business Studies*, 37(4): 525-543.
- Leibniz, G. W. (1703). Les petites perceptions. In *Nouveaux essais sur l'entendement humain*, II, chapitre 1. Paris: Garnier, 96-97.
- Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic perspectives*, 21(2), 153-174.
- Meyer, M. (2008). La méthode des scénarios : un outil d'analyse et d'expertise des formes de communication dans les organisations, *Études de communication*, 31, 2008, 133-156.
- Michel, H., Kreziak, D., & Heraud, J.-M. (2009). Evaluation de la performance des Serious Games pour l'apprentissage : Analyse du transfert de comportement des éleveurs virtuels de Vacheland. *Systèmes d'information et management*, 14(4), 71-86.
- Molinsky, A., Davenport, T. H., Iyer, B., & Davidson, C. N. (2012) Three Skills Every 21st Century Manager Needs. *Harvard Business Review*, 90(1-2), 139-143.
- Mouillot, Ph. (1999). *Fondements épistémologiques, théoriques et expérimentaux de la suggestion subliminale par le biais des images subliminales : Une approche de communication commerciale*. Thèse de doctorat, Université de Nice - Sophia-Antipolis.
- Mouillot, Ph., Drillon, D., & Montargot, N. (2018). Mosaïque des intelligences dans l'organisation : Éléments de distinction entre hauts potentiels, talents et HQI. *Revue Internationale de Psychosociologie et de Gestion des Comportements Organisationnels*, XXIV(58), 171-196.
- Pornpitakpan, C. (1999). The effects of cultural adaptation on business relationships: Americans selling to Japanese and Thais. *Journal of International Business Studies*, 30(2), 317-337.
- Roth, A., Prasnikar, V., Okuno-Fujiwara, M., & Zamir, S. (1991). Bargaining and Marketing Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study. *American Economic Review*, 81(5), 1068-1095.
- Schnell, R., Hill, P. B., & Esser, E. (2018). *Methoden der empirischen Sozialforschung*. Oldenbourg: De Gruyter.
- Schopenhauer, A. (1819). *Le monde comme volonté et comme représentation*. Traduction A. Burdeau, PUF, 1992, p. 58.
- Spearman, C. (1904). "General Intelligence", Objectively Determined and Measured, *American Journal of Psychology*, 15(1), 72-101.
- Stahl, G. K., Maznevski, M. L., Voigt, A., & Jonsen, K. (2010). Unraveling the effects of cultural diversity in teams: A meta-analysis of research on multicultural work groups. *Journal of International Business Studies*, 41(4), 690-709.
- Thomas, D. C., Elron, E., Stahl, G., Ekelund, B., Ravlin, E., Cerdin, J.-L., Poelmans, S., Brislin, R., Pekerti, A., Aycan, Z., Maznevski, M., Au, K., & Lazarova, M. (2008). Cultural Intelligence: Domain and Assessment. *International Journal of Cross-Cultural Management* 8(2), 123-143.

- Vallat, D., Bayart, C., & Bertezene, S. (2016). Serious games in favour of knowledge management and double-loop learning? *Knowledge Management Research & Practice*, 14(4), 470-477.
- Vijver, F. J. R., & Leung, K. (2009). Methodological Issues in Researching Intercultural Competence. In Deardorff, D. K. (Ed.), *The SAGE Handbook of intercultural competence*. (404- 418), Thousand Oaks, SAGE.
- Van Oudenhoven, J. P., & Van der Zee, K. I. (2002). Predicting multicultural effectiveness of international students: The Multicultural Personality Questionnaire. *International Journal of Intercultural Relations* 26(6), 679-694.
- Van Witteloostuijn, A. (2015). Toward experimental international business: Unraveling fundamental causal linkages. *Cross Cultural Management*, 22(4), 530-544.
- Zellmer-Bruhn, M., Caligiuri, P., & Thomas, D. (2016). From the Editors: Experimental designs in international business research. *Journal of International Business Studies* 47(4):,399-407.

AUTHORS

Anne Bartel-Radic

Full Professor in International Management
Univ. Grenoble Alpes, Grenoble INP, CERAG, F-38000 Grenoble, France

Philippe Mouillot

Associate Professor in Marketing & Strategy - ACR
IAE de Poitiers, France

Danielle Taylor

PhD Candidate in International Management
Univ. Grenoble Alpes, CERAG, F-38000 Grenoble, France

Apports de la Théorie Moderne de la Mesure (IRT) à l'Etude des Echelles d'Intervalle

RÉSUMÉ

Cet article présente l'intérêt de l'IRT (*Item Response Theory*), et plus particulièrement, du modèle IRT de crédit partiel adapté aux échelles d'intervalle. Il concerne la construction d'échelles en l'absence de normalité des variables de mesure et détaille les apports de l'IRT aux tests issus de la théorie classique de la mesure. Sur la base d'une collecte de données portant sur une échelle asymétrique de satisfaction, cet article décrit les étapes d'une analyse IRT de crédit partiel et les indicateurs utiles (*i.e. paramètres de discrimination et de difficulté*) pour la sélection d'items pertinents. L'IRT s'avère intéressante en complément des tests classiques en proposant un complément d'information qui pallie certains de leurs écueils.

MOTS CLÉS : IRT, théorie de la mesure, construction d'échelle, modèle IRT de crédit partiel, fiabilité de la mesure, validité de la mesure, échelle non-normale.

INTRODUCTION

En marketing, le choix de la mesure et la sélection des items revêtent un intérêt majeur. Cela conditionne en grande partie la qualité de la recherche et la crédibilité des résultats. Or, les méthodes de construction d'échelles actuellement utilisées et issues de la psychométrie classique font l'objet de nombreuses critiques (Singh, 2004). L'une des principales critiques tient au non-respect de l'hypothèse de normalité, pourtant requise comme condition d'application de l'analyse factorielle confirmatoire notamment. Les échelles développées sur la base de la théorie classique de la mesure sont aussi souvent jugées trop longues et onéreuses pour être administrées à des échantillons de non étudiants ou lors de recherches appliquées (Jong, Steenkamp et Veldkamp, 2008). Il est difficile d'identifier les items et les échelons mal compris par les répondants et donc de les éliminer ou de les reformuler. De plus, la pratique courante qui consiste à sélectionner des items à saturations élevées pour la version abrégée d'une échelle ne permet pas au chercheur de mesurer avec précision toutes les facettes d'un construit. Or, bien que la théorie classique de la mesure soit le paradigme très largement dominant pour la construction d'échelles en marketing aujourd'hui, d'autres méthodes, paramétriques et non-paramétriques, existent –plus de 50 selon Thissen et Steinberg en 1986. Parmi ses nombreuses méthodes, l'IRT (*Item Response Theory*) a l'avantage de répondre à certains des écueils de la théorie classique de la mesure, et notamment de n'imposer aucune contrainte quant à la normalité des variables de mesure (Nunnally et Bernstein, 1994).

Actuellement, pour la construction des échelles de mesure multi-items, les chercheurs en marketing, s'inspirent de la psychométrie classique (Churchill, 1989 ; Peter, 1979 ; Rossiter, 2002). Pourtant l'IRT, issu des travaux en psychométrie moderne, fait l'objet d'un intérêt constant depuis de nombreuses années dans des revues comme *Multivariate Behavioral Research* ou *Psychological Method* et est très largement diffusée en psychologie en raison de sa souplesse concernant le type de mesure employée (nominale, ordinale et intervalle) et de son apport original dans l'analyse des échelles multi-items (Thissen et Steinberg, 1988). Paradoxalement, l'IRT demeure très peu diffusée en marketing. Si dès 1985, l'article de Bechtel en vantait l'intérêt pour le marketing, relayé par l'article de Balasubramanian et Kamakura (1989), il faut attendre une dizaine d'années pour voir poindre à nouveau un intérêt pour cette méthode en marketing (De Jong et al., 2008, 2009 ; Bacon et Lenk, 2008). A l'exception de Paas et Sijtsma (2008) s'intéressant à la dimensionnalité des échelles d'intervalle, les articles publiés en marketing jusqu'à présent ne concernent que les modèles IRT adaptés aux échelles nominales.

Tous les modèles IRT suivent les mêmes grands principes mais diffèrent quant aux analyses qui peuvent être effectuées. Tous relient les caractéristiques des items de mesure et une caractéristique individuelle (appelée trait latent ou θ) à la probabilité de choisir une réponse. Ensuite, les items sont comparés sur le trait latent étudié à partir de paramètres choisis par le chercheur, e.g. *paramètre de difficulté ou de discrimination associé à chaque item*. De nombreux modèles propres à chaque type de variables (nominale, ordinale ou intervalle) et au nombre de paramètres étudiés (1, 2 voire 3 paramètres) peuvent être employés (Zhao et Hambleton, 2009). Le modèle IRT dit de crédit partiel (*Partial Credit Model* en anglais) est spécifiquement adapté à l'analyse d'échelles d'intervalle. L'analyse d'échelles via les modèles IRT offre des résultats particulièrement informatifs sur la manière dont les individus répondent aux questions (Thissen et Steinberg, 1988). L'IRT n'est pas une panacée pour pallier toutes les difficultés liées à la mesure, mais cette méthode adresse différemment les questions liées à la mesure et constitue un complément sérieux et très instructif aux tests issus de la théorie classique de la mesure (Singh, 2004).

Bien que certains auteurs comme Kline (2011) et Meade et Lautenschlager (2004) indiquent que l'IRT peut se substituer à l'analyse factorielle confirmatoire, dans cette étude, nous cherchons à montrer les apports de l'IRT –en complément des tests issus de la théorie classique de la mesure–, et ce dans le cas particulier d'une absence de normalité des variables observées. Cette absence de normalité est courante pour les échelles de mesure en marketing et peut fausser l'adéquation du modèle de mesure aux données. L'intérêt de l'IRT, et plus particulièrement du modèle IRT de crédit partiel adapté aux échelles d'intervalle, est illustré par le test d'une échelle de satisfaction client, échelle caractérisée par son absence de normalité (Babin et Griffin, 1998). La distribution de l'échelle de satisfaction client est généralement asymétrique avec un décalage à la droite de la moyenne, et donc une queue de distribution étalée vers la gauche (cas de dissymétrie négative).

Nous proposons tout d'abord une comparaison entre la théorie classique et la théorie moderne de la mesure afin de mettre en évidence les caractéristiques de l'IRT et ses apports à l'évaluation de la mesure. Puis, après une présentation générale du modèle IRT de crédit partiel, nous détaillons les étapes de l'analyse IRT en utilisant les données issues d'une échelle d'intervalle multi-items mesurant la satisfaction. Nous comparons ensuite les résultats

issus du test du modèle IRT de crédit partiel à ceux obtenus avec les tests classiques de la mesure (analyse factorielle exploratoire, analyse de fiabilité et analyse factorielle confirmatoire). Les résultats montrent que l'IRT apporte des informations complémentaires et pertinentes (discrimination, précision et difficulté) pour juger de la qualité des items d'une échelle de mesure. Elle se révèle aussi plus conservatrice que les tests classiques qui utilisés ainsi dans des conditions limitées (absence de normalité des items) conduit à conserver des items qui ne devraient pas l'être, ce qui plaide en faveur d'une complémentarité de ces méthodes.

1. APPORTS DE L'IRT A LA THEORIE DE LA MESURE

D'une manière générale, les méthodes utilisées –que ce soit la théorie classique ou moderne de la mesure– visent à proposer un nombre d'items limité tout en garantissant une qualité de mesure satisfaisante. Or toute la difficulté de la construction d'échelles tient justement au choix des items pour mesurer un construit. Nous proposons donc de rappeler les grandes étapes de la théorie classique de la mesure, puis de mettre en valeur les éléments distinctifs de la théorie moderne (IRT) incluant les problèmes soulevés par l'absence de normalité des variables de mesure. Le Tableau 1 résume les points clés de la comparaison des deux approches.

	Théorie classique de la mesure	Théorie moderne de la mesure
Utilisation	Sélection d'items et construction d'échelles	Sélection d'items et construction d'échelles
Type d'analyse	Analyse factorielle exploratoire, analyse de fiabilité, analyse factorielle confirmatoire	Analyse de la probabilité de réponse de chaque item par le modèle IRT de crédit partiel
Conditions d'application	Echelles d'intervalle avec multi-normalité des items	Echelles d'intervalle pour le modèle IRT de crédit partiel (mais pas de normalité des items requise pour l'analyse)
Nombre d'items requis pour l'analyse	Requiert un minimum de quatre items réfléchifs pour un construit réfléchif et de cinq items (deux réfléchifs et trois formatifs) pour un construit MIMIC	Requiert au minimum un item et les items de l'échelle peuvent être indifféremment réfléchifs ou formatifs
Informations sur la qualité des items	Fiabilité, validité convergente et discriminante et mesure d'adéquation aux données	Paramètres de discrimination et de difficulté associée aux passages d'échelons
Sélection des items	Choix des items sur la base de leurs inter-corrélations	Choix des items indépendamment les uns des autres sur la base de leur discrimination et de leur difficulté

Statut des items	Nécessité de spécifier <i>ex-ante</i> la nature des relations entre les items et le construit latent	Inutile de spécifier la nature des relations entre les items et le construit latent
Robustesse de l'analyse	Analyses non-robustes en cas d'absence de multi-normalité des items inclus dans l'analyse	Analyses robustes en cas d'absence de normalité de chaque item
Temps requis pour l'analyse des données	Moyen	Moyen à long lorsque les items sont corrélés à plus d'un facteur

Tableau 1 : Comparaison entre les théories classique et moderne de la mesure

1.1. La théorie classique de la mesure

La démarche psychométrique classique utilisée en marketing procède d'une analyse factorielle exploratoire (EFA, *Exploratory Factor Analysis*) dans un premier temps, suivie dans un second temps d'une analyse de fiabilité, et à titre de confirmation dans un troisième temps, d'une analyse factorielle confirmatoire (CFA, *Confirmatory Factor Analysis*). La démarche peut ainsi se résumer suivant ses trois grandes étapes.

L'analyse factorielle exploratoire permet de sélectionner les items à conserver dans l'analyse sur la base d'une étude des corrélations entre items. L'analyse en composantes principales ou en facteurs communs est généralement utilisée. Elle permet d'étudier la structure des variables incluses dans l'analyse et notamment de savoir si ces variables sont factorisables. Cette analyse permet également d'obtenir le score de chaque individu (appelé *score factoriel*) et d'estimer la valeur d'un construit comme la somme pondérée des réponses aux items de l'échelle. L'erreur de mesure est donc supposée constante pour toute valeur du construit.

L'analyse de la fiabilité d'une échelle est, quant à elle, testée via les coefficients alpha de Cronbach (1951) ou rho de Joreskog (Fornell et Larcker, 1981). Cette analyse s'appuie sur les relations entre items et notamment la corrélation *item-score total* pour évaluer l'erreur de mesure. Elle se calcule sur l'ensemble des items du construit s'il est unidimensionnel ou sur les items de chaque dimension s'il est multidimensionnel. L'analyse de fiabilité induit par son mode de calcul la sélection d'items similaires même si ceux-ci ne contribuent pas à améliorer la mesure d'autres facettes du construit. Le coefficient de fiabilité alpha ne permet pas de distinguer les items redondants de ceux qui mesurent d'autres facettes du construit ce qui peut conduire le chercheur à rechercher la maximisation de la valeur du coefficient aux dépens de la représentation réelle du construit.

L'analyse factorielle confirmatoire offre une évaluation de l'adéquation entre le modèle théorique spécifié et la réalité des données collectées. Les résultats permettent également d'évaluer la validité convergente et la validité discriminante de l'échantillon d'items. Lorsqu'un construit est étudié seul, ce type d'analyse requiert la présence d'au minimum quatre items réfléchifs pour un construit réfléchif et de deux items réfléchifs plus trois formatifs pour un construit MIMIC (Multiple Indicateurs Multiples Causes). Ce type d'analyse tend donc à

favoriser la multiplication d'items, et notamment, d'items redondants pour obtenir une meilleure validité convergente. De plus, pour utiliser l'analyse factorielle confirmatoire, la multi-normalité des variables de l'analyse est une condition d'application. Cette condition d'application des équations structurelles n'est pas toujours respectée. Or, les procédures de *bootstrap* ne permettent que partiellement de pallier le non-respect de la contrainte de multi-normalité de la mesure.

1.2. Éléments distinctifs de la théorie moderne de la mesure

La théorie moderne se distingue de la théorie classique de la mesure en termes d'information fournies sur les items, d'erreur de mesure, de statut des items, de robustesse et de lourdeur de l'analyse. Nous développons successivement chacun de ces points afin de mettre en évidence les éléments distinctifs de l'IRT et notamment du modèle IRT de crédit partiel comparé à la théorie classique de la mesure (voir Tableau 1).

1.2.1. Une évaluation plus précise de la qualité des items

Le modèle IRT de crédit partiel permet d'améliorer la construction des échelles et d'évaluer plus précisément la qualité de chacun des items (Nunnally et Bernstein, 1994). Ce modèle procure une information beaucoup plus détaillée sur la qualité de la mesure que les méthodes employées habituellement en marketing car il procure des mesures de discrimination et de difficulté des items notamment. Il permet de distinguer les items qui apportent de l'information supplémentaire des items redondants ainsi que d'identifier les items peu précis et confus ou encore les items et échelons visiblement jugés trop complexes par les répondants. Il permet ainsi de choisir des sous-ensembles d'items qui représentent bien la capacité de réponse d'une personne. En revanche, l'analyse factorielle confirmatoire est inadéquate pour détecter des items ayant des paramètres de difficultés différents (Embretson et Reise, 2000).

1.2.2. Des erreurs de mesure indépendantes entre items d'une même échelle

Avec l'IRT, l'erreur de mesure peut varier selon la valeur du construit contrairement à la psychométrie classique qui la considère constante. En analysant indépendamment chacun des items d'une échelle, l'IRT favorise la compréhension des raisons pour lesquelles la fiabilité ou la validité d'un construit ne peuvent être établies (Waller, Thompson et Wenk, 2000). Ainsi, si en psychométrie classique, la fiabilité et la validité sont calculées pour l'ensemble des items de l'échelle, avec l'IRT, les items contribuent indépendamment à la précision de la mesure, c'est-à-dire à *la variabilité de l'estimateur autour de la valeur du paramètre*. Ce principe appelé indépendance locale signifie que pour des individus qui présentent une même valeur sur le construit, les réponses aux items seront stochastiquement indépendantes.

1.2.3. Pas de nombre minimum d'items requis

Dotés de paramètres complémentaires pour évaluer la qualité d'une échelle, le chercheur utilisant l'IRT est alors en mesure d'éviter la multiplication d'items redondants dans les tests, de supprimer les items mal compris par les répondants mais aussi de reformuler les libellés d'échelons. Pour utiliser l'IRT, il n'y a pas de nombre minimum d'items requis. Cette méthode encourage donc une utilisation parcimonieuse des items et contribue à l'élaboration de tests

courts et pertinents. L'IRT permet de justifier le cas échéant l'élimination d'un item, ce que ne permettent pas les équations structurelles sans poser de problème de spécification lorsque l'échelle de mesure compte moins de quatre items réfléchifs ou deux items réfléchifs et trois items formatifs (Mackenzie et al., 2011).

1.2.4. Une absence d'hypothèse relative au statut des items

Les modèles IRT mettent en relation les probabilités de réponse des individus avec un trait latent. De ce fait, chacun des items est apprécié par sa manière de discriminer et sa difficulté et non pas par la manière dont il reflète ou forme un construit (Jarvis, et al, 2003). L'IRT est donc neutre par rapport au statut des items et aucun modèle théorique entre les items et le construit latent n'a besoin d'être spécifié ex-ante.

1.2.5. Une plus grande robustesse de la méthode

Les simulations effectuées montrent que les résultats d'une analyse factorielle confirmatoire et du modèle IRT de crédit partiel pour des échelles d'intervalle sont convergents (Kline, 2011). Par ailleurs, le modèle IRT de crédit partiel ne requiert aucune hypothèse de normalité alors que celle-ci est supposée entre les items d'un construit latent dans l'analyse factorielle confirmatoire. Ce modèle IRT est donc plus robuste qu'une analyse factorielle confirmatoire puisque la multi-normalité des items de l'analyse n'est pas une condition d'application.

1.2.6. Un temps d'analyse des données plus long

Le modèle IRT de crédit partiel est plus précis et plus robuste qu'une démarche psychométrique classique. Néanmoins, lorsque les items utilisés sont corrélés avec plus d'un facteur, l'utilisation de ce type de modèle est relativement fastidieuse. De même, son utilisation dans le cas d'échelles multi-dimensionnelles –bien que très instructive– reste assez complexe à mettre en œuvre (Paas et Sijtsman, 2008). En effet, le temps d'analyse des données est assez long dans ce cas.

2. PRESENTATION DU MODELE IRT DE CREDIT PARTIEL

Le modèle IRT de crédit partiel relie les caractéristiques des items et un trait latent (θ) à la probabilité de choisir une réponse ($P(\theta)$). Sur la base de cette variable latente, les items sont ensuite comparés entre eux en termes de discrimination, de précision et de difficulté. Le trait latent θ (ou théta) est associé à chaque item d'une échelle de mesure. Il n'a pas d'origine puisqu'il est latent et correspond à une variable d'intervalle qui permet de comparer des items ou des personnes à partir de leurs valeurs sur θ au moyen d'une transformation logistique. P est la probabilité de réponse (ou probabilité d'observation) associée à chacune des valeurs prises par le ou les traits latents du modèle. Cette probabilité varie entre 0 et 1. La fonction $P(\theta)$ liant le trait latent θ à sa probabilité de réponse est représentée par une distribution cumulée qui forme une courbe en S – typique des fonctions logistiques – et appelée courbe des caractéristiques d'un item. La probabilité de réponse tend vers 0 pour de faibles valeurs

de θ puis s'accroît seulement lorsque les valeurs de θ s'élèvent. Lorsque θ devient grand, la probabilité tend vers 1.

Pour le modèle IRT de crédit partiel, les courbes des caractéristiques d'items peuvent être exprimées de la façon suivante (Ostini et Nering, 2006) :

$$P_{i_j}(\theta) = \frac{\exp\left[\sum_{j=0}^x \alpha_i(\theta - \delta_{i_j})\right]}{\sum_{r=0}^{m_i} \left[\exp\left[\sum_{j=0}^r \alpha_i(\theta - \delta_{i_j})\right]\right]}$$

avec la contrainte :

$$\sum_{j=0}^0 \alpha_i(\theta - \delta_{i_j}) = 0$$

et pour lesquelles, l'item i est mesuré sur l'échelon $j = (0, \dots, m)$ avec $r = (0, \dots, j, \dots, m)$ et $x = (0, \dots, j)$; et où le terme α_i représente le paramètre de discrimination de l'item i et le terme δ_{ij} est le paramètre de difficulté associé à la valeur du trait latent θ de l'échelon j pour l'item i .

α (ou alpha) est mesuré par la pente de la courbe des caractéristiques d'un item à son point d'inflexion. La valeur de la pente est comprise entre 0 et plus l'infini ($+\infty$). Plus la pente est élevée, plus elle permet de différencier les répondants qui ont la capacité de répondre à l'item de ceux qui ne l'ont pas.

L'utilisation de modèles IRT de crédit partiel n'impose pas de contrainte particulière quant aux paramètres de forme et au respect de la normalité des variables de mesure de l'échelle. Cette méthode est donc adaptée aux échelles d'intervalle, et ce qu'elles soient aplaties ou pointue (kurtosis), ou bien décalées à droite ou à gauche (skewness) par rapport à une distribution normale.

3. UTILISATION DU MODELE IRT DE CREDIT PARTIEL POUR L'ANALYSE D'UNE ECHELLE DE SATISFACTION CLIENT

Le modèle IRT de crédit partiel offre la possibilité d'effectuer différentes analyses permettant de représenter, comparer et évaluer les items et leurs échelons sur le trait latent. Les étapes de l'analyse sont résumées dans le Tableau 2 qui inclut les indicateurs à analyser et leurs règles d'interprétation. Nous utilisons à titre d'illustratif l'échelle de satisfaction dont l'absence de normalité déjà observée par Babin et Griffin (1998) a été testée. Le logiciel Parscale a été utilisé pour effectuer les analyses du modèle de crédit partiel. Suivant la théorie classique de la mesure, nous avons également effectué une analyse factorielle exploratoire, une analyse de fiabilité (alpha et ρ J) ainsi qu'une analyse factorielle confirmatoire via AMOS 18 de l'échelle de satisfaction client.

Etapes	Indicateurs	Objectifs	Normes ou calcul	Interprétation
1	Paramètre de discrimination (α)	Evaluation du potentiel à discriminer d'un item	α = valeur de la pente de la courbe lorsque la probabilité $P(\theta)$ est de 50%	Eliminer un item dont le paramètre de discrimination est inférieur ou égal à 0,64 (Baker, 2001), puis sélectionner de préférence les items les plus discriminants
2	Précision d'un item (σ^2)	Evaluation de la variabilité d'un item autour de sa valeur centrale	$1/\sigma^2$ = variabilité de l'item autour de sa valeur centrale sur θ	Sélectionner de préférence les items qui présentent un faible aplatissement de leur courbe de précision, <i>c'est-à-dire privilégier les items avec une forte concentration sur la valeur centrale de θ</i>
3	Seuil de difficulté associé aux échelons d'un item (δ_{ij})	Evaluation de la qualité de représentation des échelons d'un item	δ_{ij} = valeur de θ au point de rencontre entre deux courbes de réponses successives, <i>i.e. au niveau du passage d'échelons</i>	Conserver de préférence les items dont les passages d'échelons engendrent une progression régulière sur θ

Tableau 2 : Etapes, indicateurs et interprétation d'une analyse IRT de crédit partiel

L'échelle de satisfaction utilisée comprend quatre items et un échantillon de 151 réponses a été recueilli. Ces items comprennent tous 7 échelons. Ceux-ci vont de 1 « pas du tout satisfait » à 7 « extrêmement satisfait » pour l'item 1. L'item 2 est mesuré selon le différentiateur sémantique « enchanté » - « furieux » à 7 échelons. Les items 3 et 4 sont mesurés selon des échelles de Likert de 1 « pas du tout d'accord » à 7 « tout à fait d'accord » (voir Tableau 3 pour les libellés). Les items 2 et 3 ont été inversés, *i.e. codés en sens inverse*, afin de faciliter la comparaison des items.

Items	Répartition des réponses par échelon										Asymétrie (test t)	Aplatissement (test t)
	\bar{x}	σ	1	2	3	4	5	6	7			
Item 1* – <i>Concernant mes relations avec ce commercial, je suis...</i>	5,27	1,28	2	3	13	11	47	55	20		-0,995 (-4,99)	0,943 (2,37)
Item 2** (inversé) – <i>Parlant de satisfaction à l'égard de mes relations avec ce commercial, je suis...</i>	5,26	1,18	1	4	8	15	54	52	17		-0,926 (-4,64)	1,143 (2,87)
Item 3*** (inversé) – <i>Par rapport à ce que j'attendais, je suis un peu déçu(e) par mes relations avec ce commercial</i>	5,17	1,54	4	8	7	28	27	46	31		-0,813 (-4,08)	0,070 (0,18)
Item 4*** – <i>Mes relations avec ce commercial me satisfont</i>	5,05	1,43	2	9	10	24	41	43	22		-0,695 (-3,49)	0,000 (0,00)

* 7 échelons allant de 1 « pas du tout satisfait » à 7 « extrêmement satisfait »
** différentiateur sémantique à 7 échelons allant de « enchanté » à « furieux »
*** 7 échelons allant de 1 « pas du tout d'accord » à 7 « tout à fait d'accord »

Tableau 3 : Libellés et caractéristiques de la distribution des items

L'étude de la normalité des items de l'échelle de satisfaction a été effectuée (voir Tableau 1 pour les résultats). Les items de cette échelle ont des coefficients d'assymétrie tous statistiquement significatifs au seuil de 0,05 ($t_{\text{item1}} = -4,99$; $t_{\text{item2}} = -4,64$; $t_{\text{item3}} = -4,08$; $t_{\text{item4}} = -3,49$). L'échelle de satisfaction est donc bien asymétrique et décalée vers la droite. Pour l'aplatissement, les items 1 et 2 sont plus pointus que ceux d'une distribution normale au seuil de 0,05 ($t_{\text{item1}} = 2,37$; $t_{\text{item2}} = 2,87$), alors que les items 3 et 4 sont distribués normalement ($t_{\text{item3}} = 0,18$; $t_{\text{item4}} = 0,00$). Le coefficient de multinormalité de Mardia est de 9,27 ($t = 8,22$) ce qui confirme bien l'absence de multi-normalité de l'échelle de satisfaction.

3.1. Résultats des tests modernes : l'analyse du modèle de crédit partiel

3.1.1. Etape 1 – Analyse de l'indicateur de discrimination

α (ou alpha) est un indicateur de la capacité à discriminer d'un item. Ainsi, plus la valeur de l'estimateur du paramètre α est élevée, plus l'item est jugé discriminant. Pour guider l'interprétation du paramètre α , le chercheur peut utiliser la grille suivante :

Valeurs du paramètre α	Évaluation de la discrimination d'un item
0	Nulle
$\in [0,01 ; 0,34]$	Très faible
$\in [0,35 ; 0,64]$	Faible
$\in [0,65 ; 1,34]$	Modérée
$\in [1,35 ; 1,69]$	Forte
$>1,70$	Très forte
$+\infty$	Parfaite

Extrait de F.B. Baker (2001), p 34.

Les scores de discrimination α (voir Tableau 4) montrent que l'item 2 est le plus discriminant ($\alpha_{\text{item2}} = 5,03$) alors que l'item 3 est le moins discriminant ($\alpha_{\text{item3}} = 3,77$) des quatre items. L'évaluation de la discrimination des items de l'échelle de satisfaction client est néanmoins qualifiée de très forte, selon les critères de Baker (2001).

	Estimateur de discrimination (α)	Écart type
Item 1	4,08	0,66
Item 2	5,03	0,92
Item 3	3,77	0,58
Item 4	3,80	0,62

Tableau 4 : Évaluation de la capacité des items à discriminer

3.1.2. Etape 2 – Analyse de la courbe d'information d'un item : un indicateur de précision

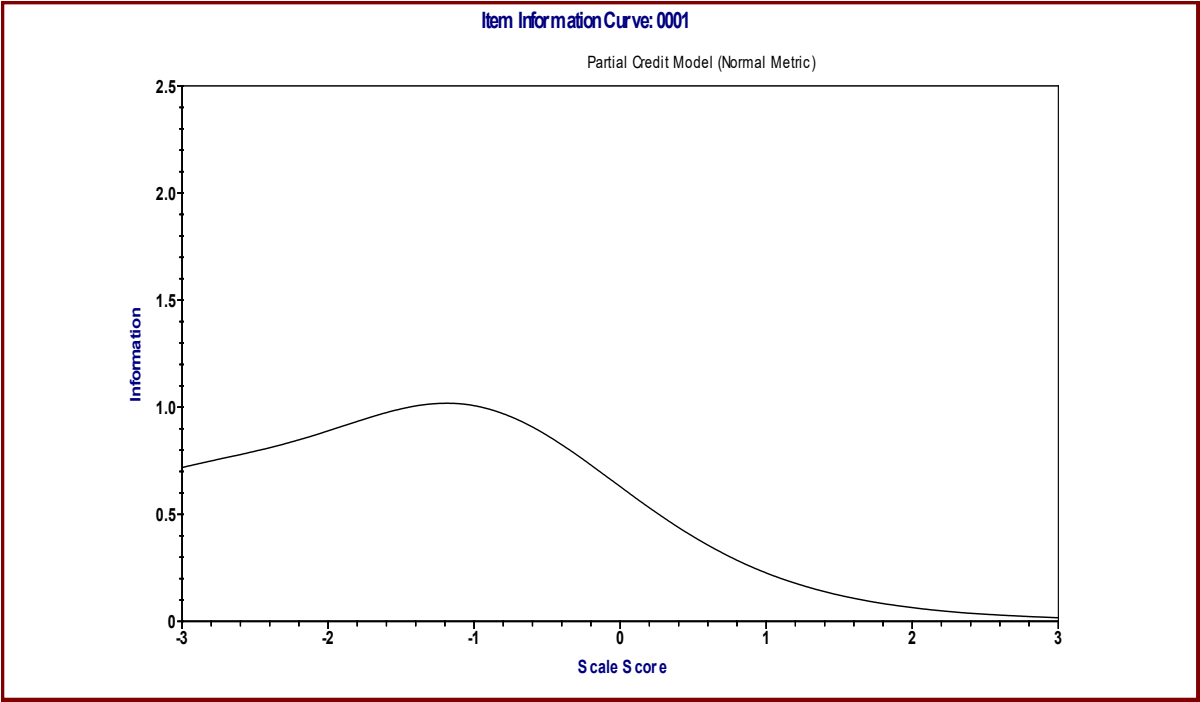
La courbe d'information d'un item propose une représentation de la précision du trait latent θ . La précision est mesurée par la variabilité de l'item autour de sa valeur centrale sur θ . La précision dépend donc de la variance de θ , dénotée σ^2 , et se calcule par la formule $1/\sigma^2$. La courbe qui en résulte présente une forme gaussienne.

La précision de l'item est grande lorsque sa variance est faible. Graphiquement plus la courbe d'information associée à un item est resserrée autour de sa valeur centrale sur θ , c'est-à-dire pointue, plus la précision de l'item est grande. Lorsque la précision de la mesure est insuffisante pour certains items, il est préférable de les supprimer de l'échelle.

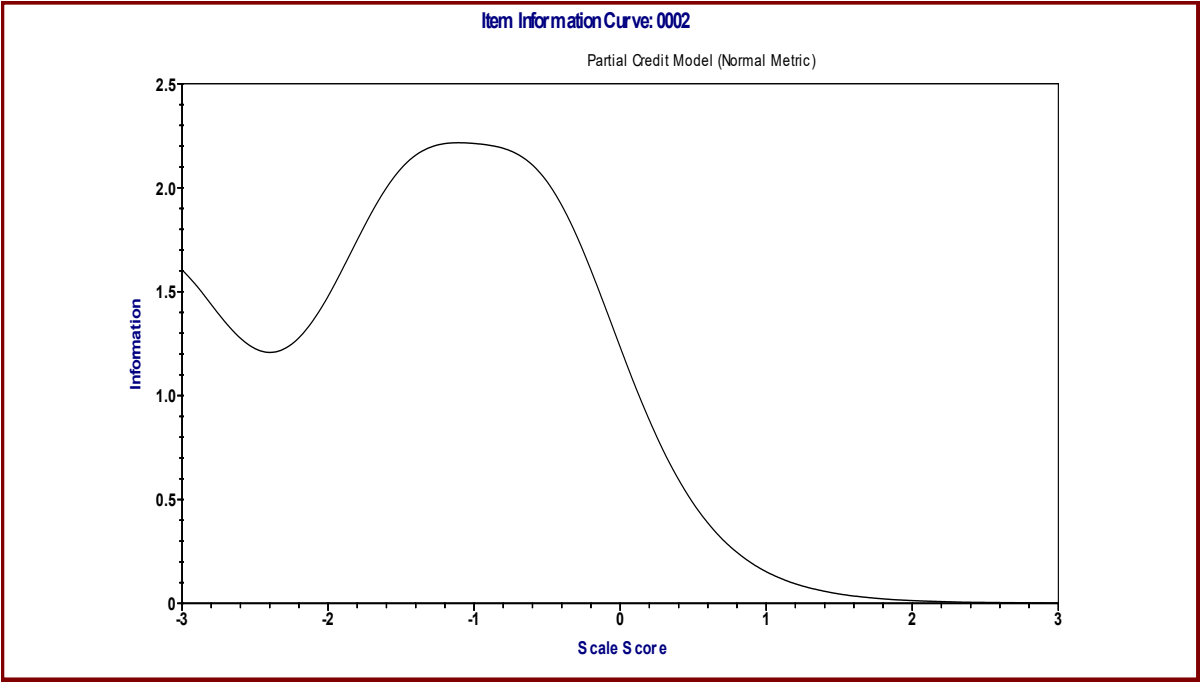
Pour l'échelle de satisfaction, les courbes d'information des items montrent que l'item 2 et dans une moindre mesure les items 1 et 4 présentent un pic au niveau de la valeur centrale θ alors que la courbe d'information associée à l'item 3 est plate (voir Figure 1). Cela signifie que les items 2, mais aussi 1 et 4 ont une précision plus grande que l'item 3 sur le trait latent étudié. L'item 3 manque donc de précision.

Figure 1 : Courbe d'information des items

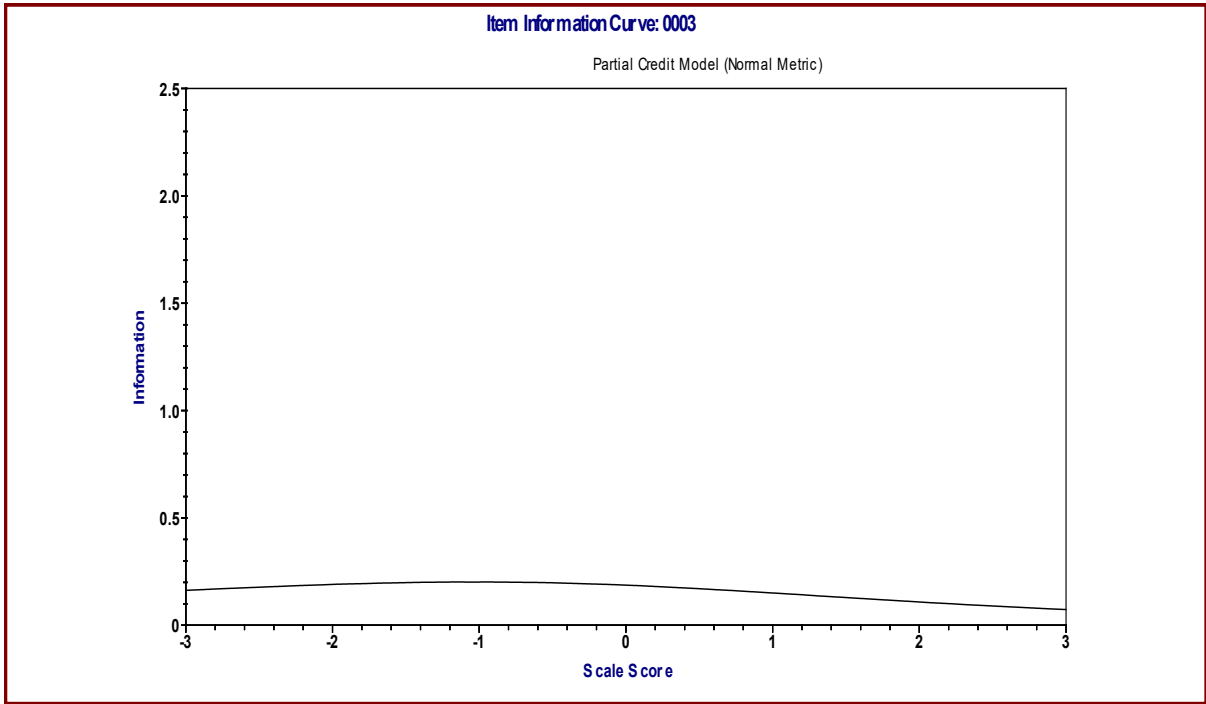
Item 1



Item 2



Item 3



Item 4

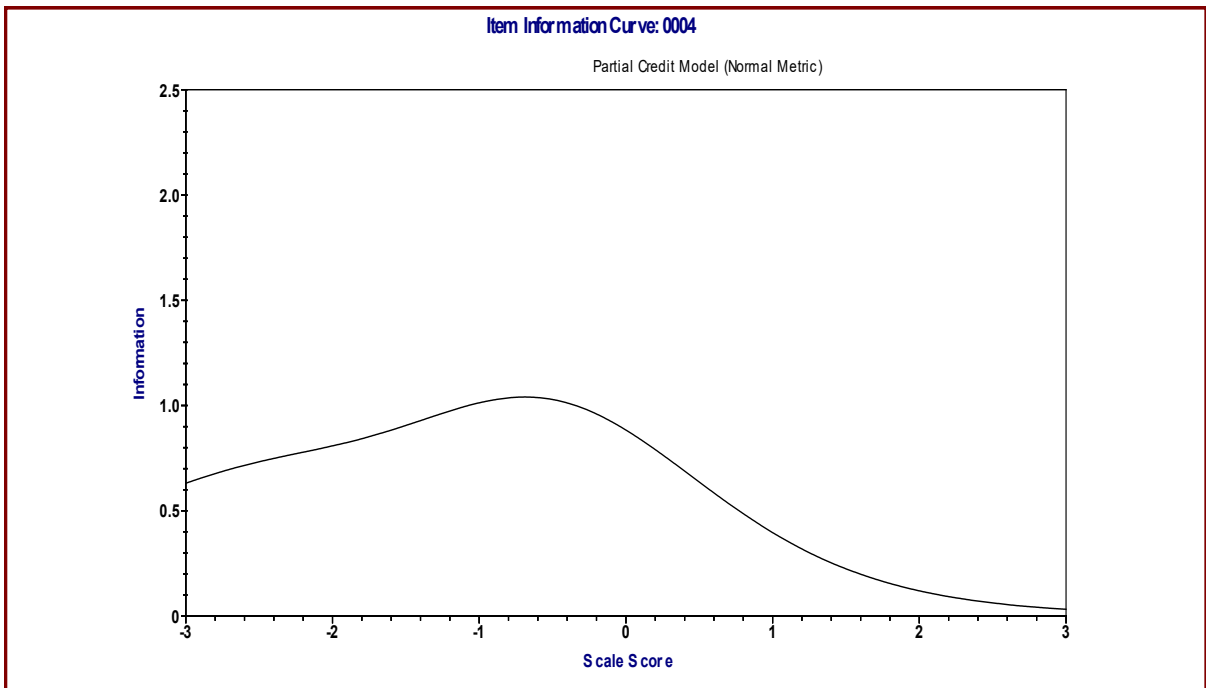
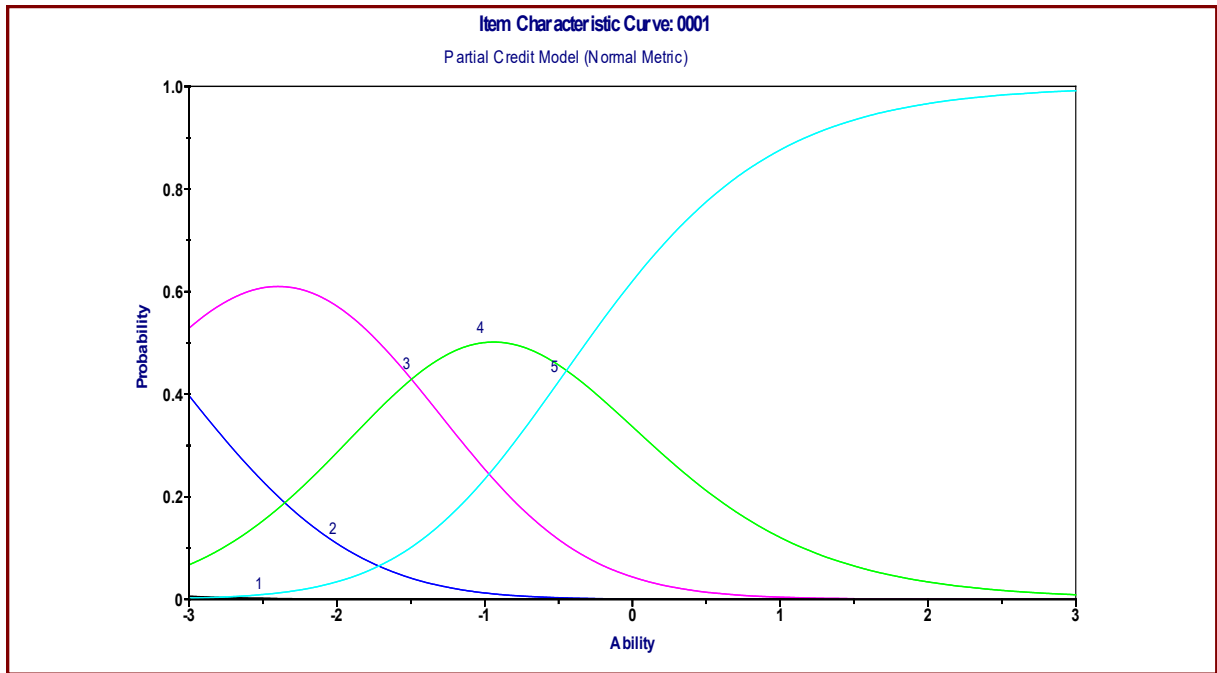
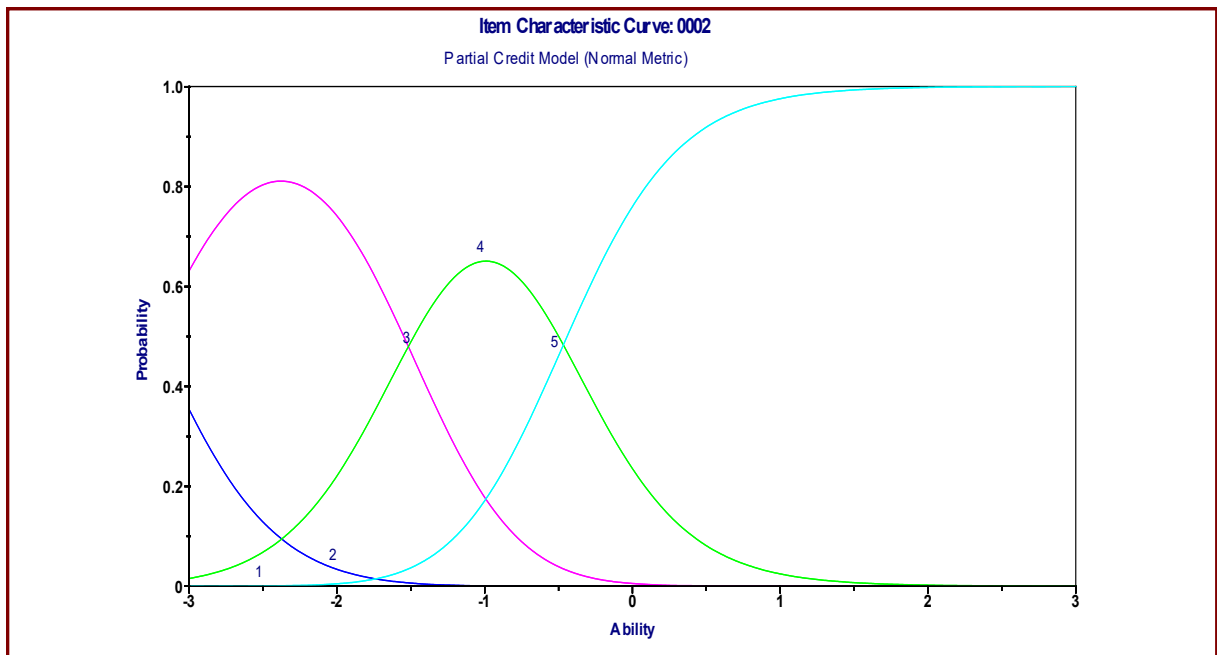


Figure 2 : Courbe des échelons des items

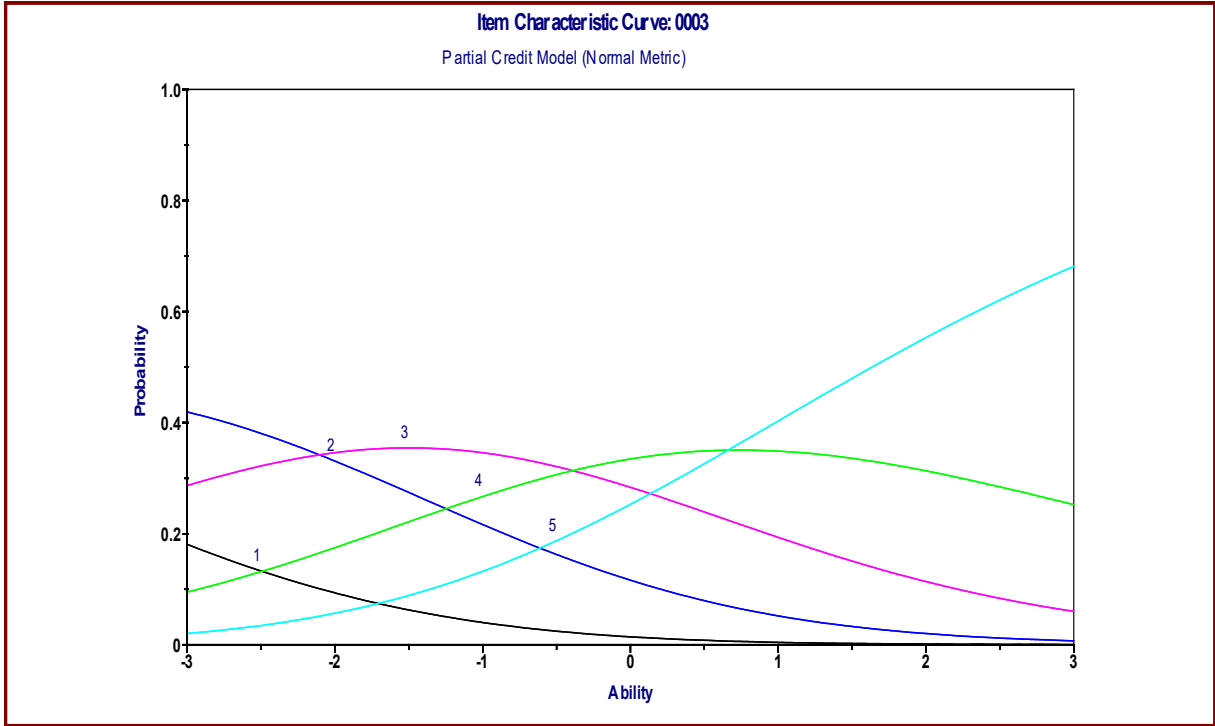
Item 1



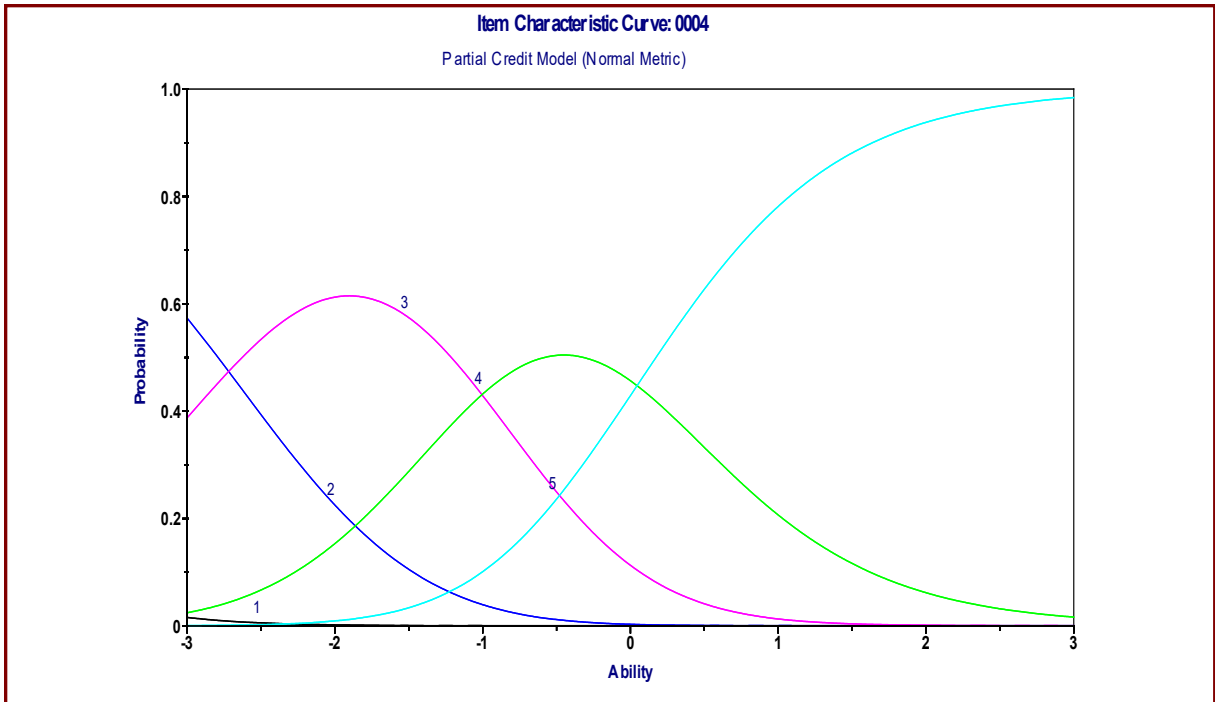
Item 2



Item 3



Item 4



3.1.3. Etape 3 : Analyse de l'indicateur de la difficulté de compréhension

δ (ou delta) est appelé paramètre de "difficulté". Il s'agit d'un indicateur de la difficulté de compréhension d'un item par les répondants. Pour des échelles multi-items et multi-échelons, la difficulté se mesure pour chaque item mais aussi pour chaque échelon sous le nom de seuil de difficulté.

Difficulté de compréhension associée aux items. L'indicateur par item correspond à la valeur de θ au point d'inflexion de la courbe des caractéristiques d'un item. Il représente le point de l'échelle de trait latent pour lequel les individus ont une probabilité de 50% de choisir l'item. La valeur δ révèle donc la plus ou moins grande difficulté de compréhension d'un item par les répondants. Ce paramètre permet de comparer les items en termes de difficulté. L'échelle de satisfaction a des indicateurs de difficulté très différents. Si les items 1, 2 et 4 ont des indicateurs de valeurs voisines ($\delta_{\text{item1}} = -2,81$; $\delta_{\text{item2}} = -2,81$; et $\delta_{\text{item4}} = -2,31$), l'indicateur de l'item 3 a une valeur beaucoup moins faible ($\delta_{\text{item3}} = -1,70$) montrant qu'il est beaucoup plus facile (voir Tableau 5).

	Estimateur de difficulté (δ)	Écart type
Item 1	-2,81	0,42
Item 2	-2,81	0,30
Item 3	-1,70	0,72
Item 4	-2,31	0,32

Tableau 5 : Evaluation de la difficulté des items

Difficulté de compréhension associée aux échelons. Les courbes de réponses aux échelons d'un item correspondent aux probabilités de réponse pour chacun des échelons d'un item et représentent la difficulté associée à ces échelons. Chaque graphique associé à un seul item présente autant de courbes que d'échelons pour l'item avec en abscisse la valeur θ associée à l'item et en ordonnée sa probabilité de réponse. Pour une échelle d'intervalle, seuls les deux échelons extrêmes présentent une courbe de réponses monotoniquement décroissante et croissante. Typiquement l'échelon qui a la plus faible valeur de difficulté parmi toutes, a une courbe de réponses qui décroît monotoniquement de 1 pour $\theta = -\infty$ à 0 pour $\theta = +\infty$. Inversement l'échelon qui a la valeur δ la plus élevée a une fonction qui s'accroît monotoniquement de 0 pour $\theta = -\infty$ à 1 pour $\theta = +\infty$. Les autres courbes sont issues de fonctions unimodales et correspondent à l'ordre des échelons. Chaque échelon correspond à un segment de la variable latente θ pour laquelle la réponse à cet échelon est probable.

L'examen des courbes de réponses aux échelons d'un item permet le calcul du seuil de difficulté associé au passage d'un échelon à un autre (δ_{ij}). Ce seuil représente la difficulté associée au passage entre deux échelons consécutifs. Notons qu'un passage d'échelon correspond au point où deux courbes de réponses consécutives se coupent. On compte donc m passages d'échelon pour une échelle à $m+1$ échelons, e.g. pour passer de « pas du tout satisfait » à « pas satisfait » nous avons un passage, puis de « pas satisfait » à « plutôt satisfait » un autre passage, soient deux passages pour trois échelons. Le seuil de difficulté

se mesure donc –en unités de θ – au point de rencontre entre deux courbes de réponses successives, *i.e. au niveau du passage d'échelons*. A chaque seuil de difficulté correspond une valeur de θ qui peut être interprétée comme la valeur pour laquelle la réponse à un échelon devient plus probable que celle fournie à l'échelon précédent. Si les intersections entre les courbes de réponses aux échelons sont bien séparées les unes des autres, cela signifie que la valeur du seuil de difficulté entre chaque échelon s'est accrue, ce qui est souhaitable (Masters, 1982). Dans notre exemple, l'échelle de satisfaction comprend 7 échelons pour chacun de ces items. Il faut donc effectuer 6 passages d'échelon pour passer de l'échelon le plus petit (1) à l'échelon le plus élevé (7) de chaque item.

L'examen de l'allure des courbes de réponses aux échelons d'un item permet, tout d'abord, d'identifier les échelons jugés difficiles pour les répondants grâce à leur plus faible probabilité à être choisis –repérable par une courbe de plus faible hauteur que les autres. Dans l'idéal, la probabilité de réponses associée à chaque passage d'échelons doit être sensiblement la même, ce qui se caractérise graphiquement par des probabilités de réponse $P(\theta)$ relativement équivalentes aux points de rencontre entre deux courbes de réponses consécutives. Enfin, la progression de la valeur du seuil de difficulté (δ_{ij}) sur la variable latente θ doit être régulière lors du passage d'un échelon à l'autre. Par progression régulière, nous entendons que la distance en valeur de θ qui sépare des seuils de difficulté (δ_{ij}) successifs doit être approximativement égale.

Dans la Figure 2, si les courbes des échelons de chaque item sont bien séparées les unes des autres. Les passages d'échelons consécutifs *i.e. intersections de courbes de réponses successives* présentent une progression régulière et constante sur la variable latente θ . Conformément aux recommandations de Masters (1982), cela montre que les échelons ont des probabilités relativement équivalentes d'être choisis sauf pour l'échelon 1 de l'item 4 qui présente une faiblesse. L'item 3 a aussi des courbes d'échelons beaucoup plus plates, ce qui confirme la faible difficulté de l'item et donc son caractère peu intéressant pour mesurer la satisfaction. En conclusion, les résultats obtenus par l'IRT montrent que l'item 3 est peu adapté à la mesure de la satisfaction.

3.2. Résultats des tests classiques

Analyse factorielle exploratoire. Les résultats de l'analyse factorielle exploratoire sont satisfaisants avec un pourcentage de variance expliquée totale de 84,5% et une qualité de la représentation de 0,85 pour l'item 1, 0,87 pour l'item 2, 0,84 pour l'item 3 et 0,82 pour l'item 4. Le test de Cattell présente un point d'inflexion très marqué à la deuxième dimension, ce qui est caractéristique d'une échelle unidimensionnelle. Les saturations ou « loadings » des items de l'échelle de satisfaction sont respectivement 0,92 pour l'item 1, 0,93 pour l'item 2, 0,92 pour l'item 3 et 0,91 pour l'item 4. Ces quatre items sont donc factorisables et semblent former une échelle de satisfaction unidimensionnelle.

Analyse de fiabilité. L'échelle de satisfaction présente une bonne fiabilité (alpha de Cronbach = 0,93 et rho de Jöreskog = 0,79). Si l'on retire de l'analyse de fiabilité chacun des quatre items successivement, le coefficient alpha à trois items reste élevé. Ainsi la corrélation des trois items avec le score total est de 0,91 sans l'item 1, de 0,91 sans l'item 2, de 0,92 sans

l'item 3 et de 0,92 sans l'item 4. Ces corrélations élevées montrent que les quatre items sont bien corrélés au construit de satisfaction (Carmines et Zeller, 1979).

Analyse factorielle confirmatoire. L'analyse factorielle confirmatoire est effectuée sur les 4 items de satisfaction, et ce malgré l'absence de normalité des items. Les résultats montrent que l'adéquation du modèle aux données est très bonne : le $\chi^2/dl = 0,80/2$ n'est pas statistiquement significatif avec $p = 0,667$, GFI = 0,997, SRMR = 0,048 et RMSEA = 0,000. Les lambdas standardisés de chacun des quatre items sont tous statistiquement significatifs : 0,902 (item 1) ; 0,913 (item 2) ; 0,883 (item 3) ; 0,805 (item 4). Le score de validité convergente (0,79) est bien supérieur au seuil de 0,5, ce qui montre que l'échelle est valide. Ces 4 items sont donc d'après les critères habituellement utilisés en marketing, un très bon reflet de la satisfaction et l'échelle mesure bien la satisfaction. Une analyse Bootstrap effectuée sur les données montre une très grande stabilité des lambdas puisque les coefficients Bootstrap sont très proches des lambdas standardisés 0,900 pour l'item 1, 0,911 pour l'item 2, 0,882 pour l'item 3 et 0,806 pour l'item 4. En conclusion, l'analyse factorielle confirmatoire montre que les 4 items choisis sont le reflet de la satisfaction.

CONCLUSION

En l'absence de normalité de la distribution d'un ou plusieurs items d'une échelle d'intervalle, l'utilisation du modèle IRT de crédit partiel est recommandée. Concernant l'échelle de satisfaction étudiée, l'observation de la distribution des items montre qu'ils sont caractérisés par une forte asymétrie. Dans ce cas, l'usage de l'IRT qui ne suppose pas de relation linéaire entre les items et le trait latent θ permet une estimation plus rigoureuse de l'échelle et de ses items. A l'issue de l'analyse IRT de l'échelle de satisfaction, seuls trois des quatre items sont conservés. L'item 3 est éliminé par manque de précision –information qui n'est pas disponible à l'issue des tests classiques. Cet item est relatif à la déception ressentie par le répondant vis-à-vis de ses relations avec le commercial avec une référence explicite aux attentes du client (« *par rapport à ce que j'attendais* »). Il ne s'agit donc pas à proprement parler d'un item mesurant la satisfaction client. Il n'est ainsi pas surprenant que l'analyse du modèle IRT mette en évidence les faiblesses de cet item. L'échelle de satisfaction comprenant trois items est ainsi préférable à celle comprenant quatre items.

L'IRT n'impose aucune condition en termes de forme pour la distribution des variables (ou items). Les tests peuvent être effectués en l'absence de normalité des variables observées. Dans les cas de variables distribuées de manière non-normale (asymétrie ou aplatissement), l'IRT peut venir en complément des tests issus de la théorie classique de la mesure pour sélectionner les items (Nunnally, 1994). L'IRT porte sur chaque item de l'échelle de mesure et ses échelons (Thissen et Steinberg, 1988), ce qui permet une analyse plus fine des items utilisés pour mesurer un construit latent que celle procurée par la méthode des équations structurelles (Reise et Widaman, 1999). L'IRT permet d'analyser de manière très détaillée la qualité des items d'une échelle de mesure. Elle affine et complète donc les analyses issues des tests classiques. Ainsi dans l'idéal, il faudrait mener successivement : (1) le test du modèle IRT pour étudier un à un les items, leurs échelons et leurs paramètres (discrimination, précision et difficulté) et sélectionner les items les plus pertinents ; puis (2) les tests classiques que sont l'analyse factorielle exploratoire pour tester la dimensionnalité du construit, l'analyse

de fiabilité pour mesurer la cohérence interne de l'échelle et l'analyse factorielle confirmatoire afin d'étudier la validité de l'échelle.

L'IRT fournit de l'information additionnelle pour juger de la pertinence des échelons d'un item en évaluant la difficulté associée à ceux-ci, ce que ne permettent pas les équations structurelles. Souvent le chercheur écarte des items pour des raisons purement liées aux procédures statistiques employées. Or certains items peuvent être pertinents mais fort mal mesurés sur le plan des échelons employés impliquant ainsi une absence de normalité. Il s'agit alors d'améliorer le libellé des échelons et leur compréhension par le répondant. Au cas par cas, l'IRT nous permet d'identifier et de sélectionner les items correctement compris par les répondants.

Globalement, l'IRT favorise la parcimonie au niveau de la construction de la mesure. Pour évaluer la qualité d'une échelle, l'IRT étudie chaque item d'une échelle de mesure via leurs paramètres de difficulté et de discrimination. Ces paramètres permettent de juger du caractère éventuellement difficile, imprécis ou redondant des items, de distinguer ceux qui représentent plus ou moins bien le construit étudié et ainsi d'éviter la prolifération d'items qui apportent peu voire très peu d'information supplémentaire pour la compréhension du construit latent. Par ailleurs, choisir les items qui ont une difficulté similaire et élevée maximise la fiabilité (Lord, 1952).

Habituellement, pour les échelles de mesure avec peu d'items le chercheur ne dispose, pour le choix des items, que du coefficient de corrélation entre items ou lorsque le nombre d'items est égal à 3 de la valeur du coefficient de Cronbach qui permet d'étudier la fiabilité de l'échelle. L'IRT permet quant à elle l'analyse de la variable latente exprimée par la probabilité de réponse à un item et à chaque échelon. Elle apporte des informations nouvelles sur le pouvoir discriminant d'un item pour sa variable latente, la précision d'un item et la difficulté associée au passage d'échelon au sein de chaque item. Elle permet donc le choix des items les plus représentatifs de la variable latente, et cela quel que soit la normalité de la distribution des items.

L'IRT répond particulièrement bien à des lacunes de l'analyse factorielle confirmatoire en offrant la possibilité d'évaluer la pertinence de chacun des items et la qualité des échelons. L'IRT permet ainsi d'éviter la multiplication d'items redondants dans les tests, d'identifier les items mal compris et les libellés d'échelons mal formulés. L'IRT est donc à utiliser en complément de l'analyse factorielle confirmatoire pour s'assurer que les items et échelons sont bien spécifiés et compris des répondants ou pour justifier l'élimination de ceux qui ne le seraient pas, voire même, en substitut de l'analyse factorielle confirmatoire lorsque que celle-ci ne peut être utilisée (*e.g. échelles de moins de quatre items*). L'utilisation de l'IRT pour l'étude d'échelles d'intervalle unidimensionnelles est préconisée par Meade et Lautenschlager (2004) en raison des informations complémentaires et utiles à la prise de décision des managers et des chercheurs. Par contre, l'IRT est plus complexe d'utilisation que l'analyse factorielle confirmatoire pour le test de construits multidimensionnels ou le test de modèle de mesure. Notons néanmoins que les travaux récents de Paas et Sitjman (2008) sont prometteurs et devrait bientôt donner lieu à l'élaboration d'un logiciel de test de modèle IRT de crédit partiel pour le test d'échelles multidimensionnelles.

Le frein principal en vue d'une généralisation de l'application de l'IRT en marketing est surtout le manque d'ergonomie des logiciels. Au-delà de leur intérêt scientifique, leur mise en pratique peut être quelque peu complexe et aucune procédure d'analyse n'est actuellement disponible sur les logiciels classiques en marketing tels qu'IBM-SPSS, SPAD ou Statistica. Le

développement logiciel constitue donc une perspective d'amélioration majeure qui favoriserait l'utilisation de cette méthode en marketing.

REFERENCES

- Babin B. L. et Griffin M. (1998), The Nature of Satisfaction: An Updated Examination and Analysis, *Journal of Business Research*, 41, 2, 127-136.
- Bacon L. et Lenk P. (2008), Breaking the binary model code: IRT models offer flexible marketing measurement, *Marketing Research*, 20, 4, 7-10.
- Balasubramanian S. K. et Kamakura W. (1989), Measuring consumer attitudes toward the marketplace with tailored interviews, *Journal of Marketing Research*, 26, 3, 311-326.
- Baker F. B. (2001), *The basics of item response theory*, 2nd édition, University of Maryland, College Park, MD., ERIC Clearinghouse on Assessment and Evaluation.
- Bechtel G. (1985), Generalizing the Rasch model for consumer rating scales, *Marketing Science*, 4, 1, 62-75.
- Bergkvist L. et Rossiter J. R. (2007), The predictive validity of multiple items versus single item measures of the same constructs, *Journal of Marketing Research*, 44, 2, 175-184.
- Bergkvist L. et Rossiter J. R. (2009), Tailor-made single-item measures of doubly concrete constructs, *International Journal of Advertising*, 28, 4, 607-621.
- Cacioppo J. T. et Petty R. E. (1982), The need for cognition, *Journal of Personality and Social Psychology*, 42, 1, 116-131.
- Carmines E. et Zeller R. (1979), *Reliability and validity assessment*, Quantitative Application in the Social Sciences, Sage publications.
- Chernyshenko O. S., Stark S., Chan K-Y, Drasgow F. et Williams B. (2001), Fitting item response theory models to two personality inventories; issues and insights, *Multivariate Behavioral Research*, 36, 4, 523-562.
- Churchill G. (1979), A paradigm for developing better measures of marketing constructs, *Journal of Marketing Research*, 16, 1, 64-73.
- Cronbach L. J. (1951), Coefficients Alpha and the internal structure of tests, *Psychometrika*, 16, 3, 297-234.
- De Jong M. J., Steenkamp J., Fox J. et Baumgartner H. (2008), Using item response theory to measure extreme response style in marketing research: a global investigation, *Journal of Marketing Research*, 45, 1, 104-115.
- De Jong M. J., Steenkamp J. et Veldkamp B. P. (2009), A model for the construction of country-specific yet internationally comparable short-form marketing scales, *Marketing Science*, 28, 4, 674-689.
- Embretson S. E. et Reise S. P. (2000), *Item response theory for psychologists*, London, Lawrence Erlbaum Associates.
- Jarvis C. B., MacKenzie S. B. et Podsakoff P. M. (2003), A critical review of construct indicators and measurement model misspecification in marketing and consumer research, *Journal of Consumer Research*, 30, 2, 199-218.
- Jolibert A. et Jourdan P. (2006), *Marketing research : méthodes de recherche et d'études en marketing*, Paris, Dunod.
- Kline R. B. (2011), *Principles and practice of structural equation modeling*, 3^{ème} édition, London, Guilford Press.

Lord F. (1952), The relationship of the reliability of multiple choice items to the distribution of items difficulties, *Psychometrika*, 18, 181-194.

MacKensie S., Podsakoff P. M. et Podsakoff N. P. (2011), Construct measurement and validation procedures in MIS and behavioral research: integrating new and existing techniques, *MIS Quarterly*, 35, 2, 293-334.

Masters G. (1982), A Rasch model for partial credit scoring, *Psychometrika*, 47, 2, 149-174.

Meade A. W. et Lautenschlager G. J. (2004), A comparison of item response theory and confirmatory factor analysis for establishing measurement equivalence/invariance, *Organizational Research Methods*, 7, 361-388.

Nunnally J. C., Bernstein I. H. (1994), *Psychometric theory*, 3ème édition, New York, Mc Graw Hill.

Ostini R., Nering M. (2006), *Polytomous item response theory models*, London, Sage.

Paas L., Sijtsma K. (2008), Nonparametric item response theory for investigating dimensionality of marketing scales: a SERVQUAL application, *Marketing Letters*, 19, 2, 157-170.

Peter J. P. (1979), Reliability: a review of psychometrics basics and recent marketing practices, *Journal of Marketing Research*, 16, 1, 6-17.

Reise S., Widaman K. (1999), Assessing the fit of measurement models at the individual level: a comparison of item responses theory and covariance structure approaches, *Psychological Methods*, 4, 1, 3-21.

Rossiter J. R. (2002), The C-OAR-SE procedure for scale development in marketing, *International Journal of Research in Marketing*, 19, 4, 305-335.

Singh J. (2004), Tackling measurement problems with item response theory; principles, characteristics and assessment with an illustrative example, *Journal of Business Research*, 57, 2, 184-208.

Slijtsma K., Hemker B. (1998), Non-parametric polytomous IRT models for invariant item ordering with results for parametric models, *Psychometrika*, 63, 2, 183-200.

Thissen D., Steinberg L. (1988), Data analysis using item response theory, *Psychological Bulletin*, 14, 3, 385-395.

Vrignaud P. (1996), Les tests du 21 Siècle : que peut-on attendre des évolutions méthodologiques et technologiques dans la classification et l'évaluation des personnes?, *Pratiques Psychologiques*, 2, 5-28.

Waller N., Thompson J. et Wenk E. (2000), Using IRT to separate measurement bias to group differences on homogeneous and heterogeneous scales: an illustration with the MMPI, *Psychological Methods*, 5, 1, 125-146.

Zhao Y., Hambleton R. (2009), *Software for IRT Analyses: Description and Features*, Center for Educational Assessment Research, Report N° 652, Amherst, MA: University of Massachusetts.

AUTEURS

Ping Lei

Professeur assistant en marketing
INSEEC School of Business & Economics, F-69007 Lyon, France

Maud Dampérat

Professeur des universités en marketing
Univ Lyon, UJM-Saint-Etienne, COACTIS, EA 4161, F-42023, Saint-Etienne, France

Florence Jeannot

Professeur associé en marketing
INSEEC School of Business & Economics, CERAG, EA 7521, F-69007 Lyon, France

Alain Jolibert

Professeur émérite en marketing
INSEEC School of Business & Economics, CERAG, EA 7521, F-69007 Lyon, France

IRT Approach for Rating Scales: Applications for Normal and Non-Normal Distributions

ABSTRACT

In administrative sciences, one of the main challenges is to choose the right items for a measurement scale. The purpose of this article is to provide marketing researchers a detailed description of item response theory (IRT) for rating scales. It details the different stages of IRT using the graded response model (GRM) on two rating scales (need for cognition and satisfaction). The IRT approach offers a notable advantage due to its ability to precisely assess the quality and the contribution of each of the items to the latent trait. GRM could be used either as a complement or a substitute to the confirmatory factor analysis (CFA), especially for non-normal distributed scales.

KEY WORDS: Item Response Theory (IRT), Graded Response Model (GRM), Rating Scale, Classical Test Theory (CCT), Measurement Theory

INTRODUCTION

In marketing, the choice of measurement and selection of items are of major interest because it largely determines the quality of research and the credibility of the results. However, multi-item scale development methods that are currently used and inherited from classical psychometrics are the subject of much criticism. Multi-item scales are often considered too long and expensive to be administered to non-student samples or in applied studies (Jong, Steenkamp & Veldkamp, 2008). It is also difficult to identify items and rating points scales that are misunderstood by respondents and thus need to be eliminated or reformulated. In addition, the common practice to select items with high loadings for a short version of a scale does not allow the researcher to accurately measure all aspects of a construct. As a result, an investigation of the methods and practices used in modern psychometric to build scales and their interest for marketing is necessary.

Currently, for building multi-item scales, marketing researchers are often inspired by classical test theory (CCT) (Churchill, 1989; Peter, 1979; Rossiter, 2002) using several methods to select items to measure a construct and to study their reliability and validity. The most commonly used approach in marketing can be summarized in three steps. The first step relies upon an exploratory factor analysis (principal component factor analysis or common factors analysis) and to select items based on a study of how they are correlated with each other. This analysis is also used to obtain a score for each individual (called "factor score") and estimate the value of a construct as a weighted sum of the responses to the items of the scale. The

measurement error is assumed to be constant for all values of the construct. In the second step, the reliability of the scale is tested using Cronbach's alpha coefficient (1951) or Joreskog's rho (Fornell & Larcker, 1981). The reliability analysis is based on the relationships between items, especially the total item-score correlation of the test to evaluate the measurement error. The reliability is computed on all the items if the scale is unidimensional or on the items of each dimension if the scale is multidimensional. The reliability analysis favors the selection of similar items. The alpha coefficient does not distinguish the redundant items from the original ones. Its only purpose is the maximization of the value of the coefficient. For the third step, convergent validity and discriminant validity of the sample items are evaluated using a confirmatory factor analysis (CFA). However, structural equation modeling tends to favor the multiplication of items (at least four items when one construct is studied), especially redundant ones, to get a better validity. Thus, the difficulties of developing scales are mainly related to choosing the right items to measure a given construct and can be summarized as follows: how to reduce the number of items used while ensuring a satisfactory quality of measurement?

To improve the scales construction, modern psychometrics now rely mainly on the item response theory (IRT) approach (Nunnally & Bernstein, 1994). It provides much more detailed information on the quality of the measurement than the methods commonly used in marketing, including measures of discrimination, accuracy and difficulty of items. IRT approach distinguishes items by providing additional information about redundant items, identifies inaccurate and confusing items or rating points obviously considered too complex by respondents. With this additional information to assess the quality of a scale, the researcher is then able to avoid the proliferation of redundant items and remove items misunderstood by respondents. This favors the development of short and relevant scales. IRT assesses the quality of each item more accurately (Nunnally & Bernstein, 1994; Singh, 2004) and justifies the item disposal.

IRT differs from CCT in the sense that measurement error may vary depending on the value of the construct while CCT considers it as a constant. Independently analyzing each item of a scale, IRT makes it easier to understand why the reliability or validity of a construct cannot be established (Waller, Thompson, & Wenk, 2000). Thus, if with the CCT approach, reliability and validity are calculated for all the items of a scale, with the IRT approach, the items independently contribute to the accuracy of the measurement. This principle called "local independence" means that for people obtaining the same value on a construct, responses to items are stochastically independent. IRT has been of constant interest for many years in journals such as *Multivariate Behavioral Research* and *Psychological Method* and is widely used in psychology because of its flexibility according to the type of measurement used (binary, ordinal or rating scales) and its original contribution to the analysis of multi-item scales (Thissen & Steinberg, 1988).

Paradoxically, IRT remains rarely used in marketing. In 1985, Bechtel boasted its interest in marketing, backed by the article of Balasubramanian and Kamakura (1989). It took a decade to see the emergence of a renewed interest for this method (De Jong et al., 2008, 2009; Lenk & Bacon, 2008), especially thanks to the contribution of Singh (2004). Yet its interest is obvious

considering the issues raised by the selection of items during the development/use of a scale. Singh (2004), the most complete article on IRT, uses rating scales measuring role conflict and role ambiguity concepts. To show the interest of the IRT approach, these two rating scales have been converted into dichotomous scales, which was a prerequisite for using Multilog software. At that time, rating scales could not be analyzed as such because the available software was of great complexity (Singh, 2004). At the present time, various software such as Mplus and Parscale allow to easily examine rating scales using an IRT approach.

The purpose of this article is precisely to present the IRT approach as applied to rating scales (Malhotra & Birks, 2006). The specific application of IRT approach for rating scales is called graded response model (GRM). GRM generates simple graphical curves to select the appropriate items and is proposed as a complement to the CCT, but also as an alternative to it, especially when the scales comprise less than four items. The originality of this article is the proposal of a three-step approach including indicators and interpretation rules for the application of IRT to rating scales in marketing. To demonstrate GRM's potential interest in marketing, we used two scales: one is normally distributed (need of cognition) while the other follows a non-normal distribution (customer satisfaction). And we compare the results from GRM analysis to those obtained with CCT.

1. BASIC ELEMENTS OF THE IRT METHOD

IRT models, also called "latent trait models," follow the same principles but differ in the analysis that can be performed. They connect all item characteristics and an individual characteristic (latent trait θ) to the probability of choosing a response. Then items can be directly compared to the trait being studied based upon parameters selected by the researcher (e.g., the difficulty and discrimination associated with each item parameter).

1.1. Value of the latent variable associated with an item (θ)

The latent trait associated to each item of a scale is called θ (or theta). It represents the person's ability to respond to an item. θ has no origin as it is a rating latent variable that allows to compare items or people from their values on θ . It is constructed to have a mean of 0 and a variance of 1. θ is the latent trait used in the computation of all IRT models.

1.2. Response probability values of a latent variable (P)

P is the probability to answer an item "right." This probability varies between 0 and 1. The function $P(\theta)$ linking the latent variable θ to the probability of response is represented by a cumulative distribution which forms an S-curve (typical of logistic functions) called item characteristic curve (ICC). Response probability tends to 0 for low values of θ and then increases only when the values of θ increase. When θ becomes high, the probability tends to 1.

1.3. Evaluation of the difficulty of an item (δ)

δ (or delta) is called a "difficulty" parameter. It is an indicator of the difficulty of understanding an item by respondents. This parameter corresponds to the value of θ at the inflection point of the item characteristic curve. It represents the point on the latent trait scale for which individuals have a 50% probability of choosing the item. Items with negative values of δ are considered easier because people are relatively more likely to choose them, while items with positive values of δ are considered more difficult. The value of δ thus indicates the greater or lesser difficulty of understanding an item by respondents. This parameter allows the comparison of the difficulty between the items. The simplest IRT model is the one-parameter IRT model, which is also called the Rasch model (Bechtel, 1985). It only evaluates the difficulty parameter (δ) using binary data such as Boolean-type measured from answers to yes/no questions. The one parameter IRT model function is as follows:

$$P(\theta) = 1 - 1 / (1 + e^{-(\theta - \delta)})$$

δ parameter is also used in more complex IRT models in addition to other indicators of accuracy of the item. For multi-item and multipoint scales, the difficulty is measured for each item and for each point under the name of threshold difficulty.

1.4. Assessment of the discrimination of an item (α)

α (alpha) parameter is called "discrimination." This parameter is measured by the slope of the characteristics of an item to its inflection point. The slope value is between 0 and infinity ($+\infty$). The steeper the slope, the higher it can differentiate respondents who have the ability to respond to an item "right" from those who have not. To evaluate the discrimination of an item, the two-parameter IRT models are needed because two-parameter IRT models compute not only the difficulty of the item (δ) but also item discrimination (α). In this way, items can be differentiated from low to high responses by each respondent. For a model with two parameters (α and δ), the function becomes:

$$P(\theta) = 1 - 1 / (1 + e^{-\alpha(\theta - \delta)})$$

These two-parameter models are used for all types of dichotomous scales.

2. IRT MODEL AND INDICATORS FOR RATING SCALES

Different models of IRT have been adapted to variable types (binary, ordinal or rating), and the number of parameters selected. These models can be estimated using software such as Multilog, Mplus, Rumm, Bilog, Parscale or using different mathematical principles according to the number of parameters to be estimated and the type of measurement scale used. Following the presentation of the GRM with two parameters (α and δ), whose specifications are appropriate for rating scales, we successively study the different indicators available for this type of model and their interpretation rules.

2.1. Graded response model suitable for rating scales

IRT models with two-parameters (α and δ) have been adapted to the measurement scales commonly used in marketing such as attitude scales, personality or involvement. The GRM, as an extension of the two-parameter model, is the IRT model suited to rating scales such as Likert scales (Samejima, 1969). It is also called the Rasch model with multiple responses or rating scale (Sijtsman & Hemker, 1998). In GRM, each scale item is described by one item slope parameter α and j from 1 to m between category threshold parameter δ_{ij} . For GRM, the characteristic curves of items can be expressed based on Samejima (1969):

$$P_{ij}(\theta) = \frac{e^{[\alpha i(\theta - \delta_{ij})]}}{1 + e^{[\alpha i(\theta - \delta_{ij})]}}$$

where $j = 1, \dots, m$, j being the between category threshold parameter, and i the item scale.

2.2. Indicators and interpretation rules for the use of the GRM

GRM provides the opportunity to perform various analyses to represent, compare and evaluate the items and their rating points on the latent trait θ . The steps of the analysis are summarized in Table 1. At each stage, the indicators to analyze and their rules of interpretation are specified.

Steps	Indicators	Analysis	Objectives	Computational norms	Interpretation
1	Discrimination parameter of an item (α)	Analysis of the shape of an item characteristics curve (S logistic curve)	To test how well does the item differentiate low from high responses	α = slope value for the probability $P(\theta)$ equal to 50%	Remove items whose discrimination parameter is less than or equal to 0,64 (Baker, 2001), and selecting preferentially the most discriminating items
2	Accuracy of an item (σ^2)	Analysis of the shape of the information curve of an item (bell curve of a normal distribution)	To test the variability of an item around its central value	$1/\sigma^2$ = variability of an item around its central value on θ	Selecting preferentially items having a positive kurtosis on θ
3	Difficulty threshold of each rating point	Analysis of the response curves to each rating point of an item (bell	To test how difficult an item can be understood	δ_{ij} = value of θ at the intersection point between two successive	Selecting preferentially items having successive rating points showing a steady increase on θ

of an item (δ_{ij})	curve of a normal distribution)	by respondents	response curves	and whose responses probability $P(\delta_{ij})$ are equivalent between rating points
---------------------------------	---------------------------------------	-------------------	--------------------	--

Table 1: Steps, indicators and interpretation of a rating scale IRT

The steps in the analysis and interpretation of indicators of IRT are emphasized in using a shortened version of the need for cognition (NFC) scale. The NFC was administered to a sample of 184 consumers. The scale consists of five items from the Cacioppo and Petty (1982) scale. These items are rated on a five-point Likert scale varying from 1 "strongly agree" to 5 "strongly disagree." The wording of items and the number of occurrences per point per item are provided in Table 2.

Items*	σ	1	2	3	4	5	Skewness (test t)	Kurtosis (test t)	
NFC1 - I am not satisfied unless I am thinking.	2.71	1.30	31	63	22	23	45	0.217 (1,20)	-0.917 (-2,54)
NFC2 - I would rather do something that requires little thought than something that is sure to challenge my thinking abilities	2.57	1.22	42	48	62	12	20	0.479 (2.65)	-0.503 (-1.39)
NFC3 - I only think as hard as I have to.	2.06	1.30	92	32	32	13	15	0.976 (5.40)	-0.237 (-0.65)
NFC4 - I prefer to think about small, daily projects to long-term ones.	2.37	1.29	66	35	48	19	16	0.535 (2.96)	-0.800 (-2.21)
NFC5 - Learning new ways to think doesn't excite me very much.	2.18	1.24	73	46	41	10	15	0.853 (4.72)	-0.201 (-0.55)

* 5-point scales from 1 "strongly agree" to 5 "strongly disagree"

Table 2: NFC scale – Items distribution of point scores for each item

This scale has acceptable reliability (Cronbach's alpha = 0.617 and Jöreskog rho = 0.626). The correlations between each of the five items and the total score varies for items NFC1 and NFC3 (respectively 0.551 and 0.577) to a value greater than 0.6 for items NFC2 (0.662), NFC4 (0.672) and NFC5 (0.667), showing that these items are correlated to the construct (Carmines & Zeller, 1979). A confirmatory factor analysis of the data indicates that the model fit is satisfactory: Chi ²/df = 6.5/5 is not statistically significant with p = 0.257, GFI = 0.986, SRMR = 0.036 and RMSEA = 0.041. The standardized Lambdas of each of the five items are all

statistically significant: 0.328 (NFC1), 0.562 (NFC2), 0.416 (NFC3), 0.601 (NFC4), 0.584 (NFC5). The convergent validity index (0.260) is less than the threshold of 0.50, showing a problem within the scale. Generally, it is advised to suppress the items with the lower loading, NFC1 and NFC3.

2.1.1. Step 1 - Item characteristic curve: an indicator of discrimination

The characteristic curve of an item is used to evaluate the discriminating power of an item for a given latent trait. Each curve is a function which relates values of a latent variable (θ) with its observation probability ($P(\theta)$). For a multi-item scale, there are as many curves as the number of items in the scale. These curves show a logistic distribution linking the probability of response as ordinate values to θ the latent variable as abscissa. These characteristic curves are particularly important because they are used not only to study the properties of those items but also those of the latent trait they measure. The juxtaposition of the curves of each item allows the evaluation of the different capacities of the items in the representation of the construct and thus their comparison.

The examination of the curves is used to evaluate the relevance of an item and eliminate those that are not discriminatory (Nunnally & Bernstein, 1994; Vrignaud, 2002). Thus an item whose shape characteristics curve deviates too much from the logistics perfect shape (S-curve), especially if the distribution is too flat, is eliminated. Such flattened curves mean that the item is not discriminatory and therefore provides no additional information on the scale (Jolibert & Jourdan, 2011). After the examination of the curves' shape, the study of discrimination parameters can be added. These parameters present the same information as the curves shapes examination. The higher the value of the α parameter, the higher the item will be considered discriminatory. To guide the interpretation of the α parameter, the researcher can use the following table (Table 3).

Value of the parameter α	Evaluation of the item discrimination
0	Nul
$\in [0,01 ; 0,34]$	Very low
$\in [0,35 ; 0,64]$	Low
$\in [0,65 ; 1,34]$	Moderate
$\in [1,35 ; 1,69]$	Strong
$>1,70$	Very Strong
$+\infty$	Perfect

Table 3: Criteria of discrimination parameters values. Source Baker (2001), p. 34.

Rule of Interpretation: Following a review of the curves and/or study parameters of discrimination (α), any item whose distribution is too flat or that the estimator has a discrimination qualified low to zero ($\alpha \leq 0.64$) must be eliminated as the item is found non-discriminating. It is advised to select the most discriminating items.

Illustration with need for cognition scale: Mplus software was used to perform the GRM analyses. The characteristic curves of the items show that the items NFC2, NFC4 and NFC5 differ from the other two (NFC1 and NFC3) (see Figure 1). The curves associated with NFC1 and NFC3 show flatter slopes than the curves of NFC2, NFC4, and NFC5 indicating a lower degree of discrimination.

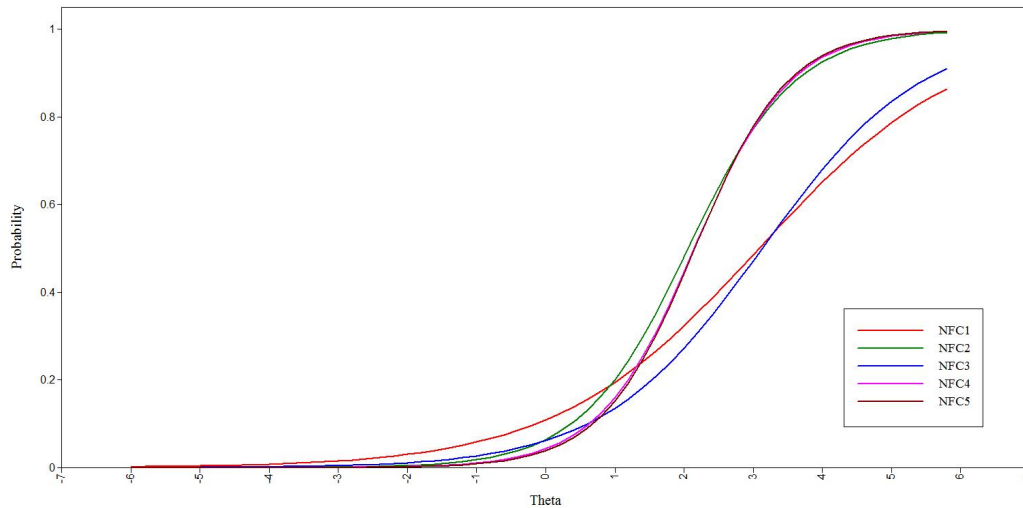


Figure 1: Characteristic curve for the 5 NFC items

Table 4 also shows that the three curves associated with NFC2, NFC4 and NFC5 are characterized by discrimination estimators' values α qualified as strong according to Baker's criteria (2001): respectively 1.301, 1.446 and 1.489. Discrimination estimators' values for items NFC1 and NFC3 show a moderate discrimination, 0.681 and 0.868 respectively. Items NFC1 and NFC3 provide less information about the latent trait of need for cognition than the other three, especially the item NFC1 which is just above the threshold of 0.64 set by Baker (2001). For further analysis and for the sake of simplicity, we chose to compare the items NFC1 and NFC3 with the NFC5 item, because items NFC1 and NFC3 have very different characteristics curves from items NFC2, NFC4 and NFC5.

Items	Discrimination parameters (α)	SD
NFC1	0.681	0.242
NFC2	1.301	0.333
NFC3	0.868	0.255
NFC4	1.446	0.359
NFC5	1.489	0.399

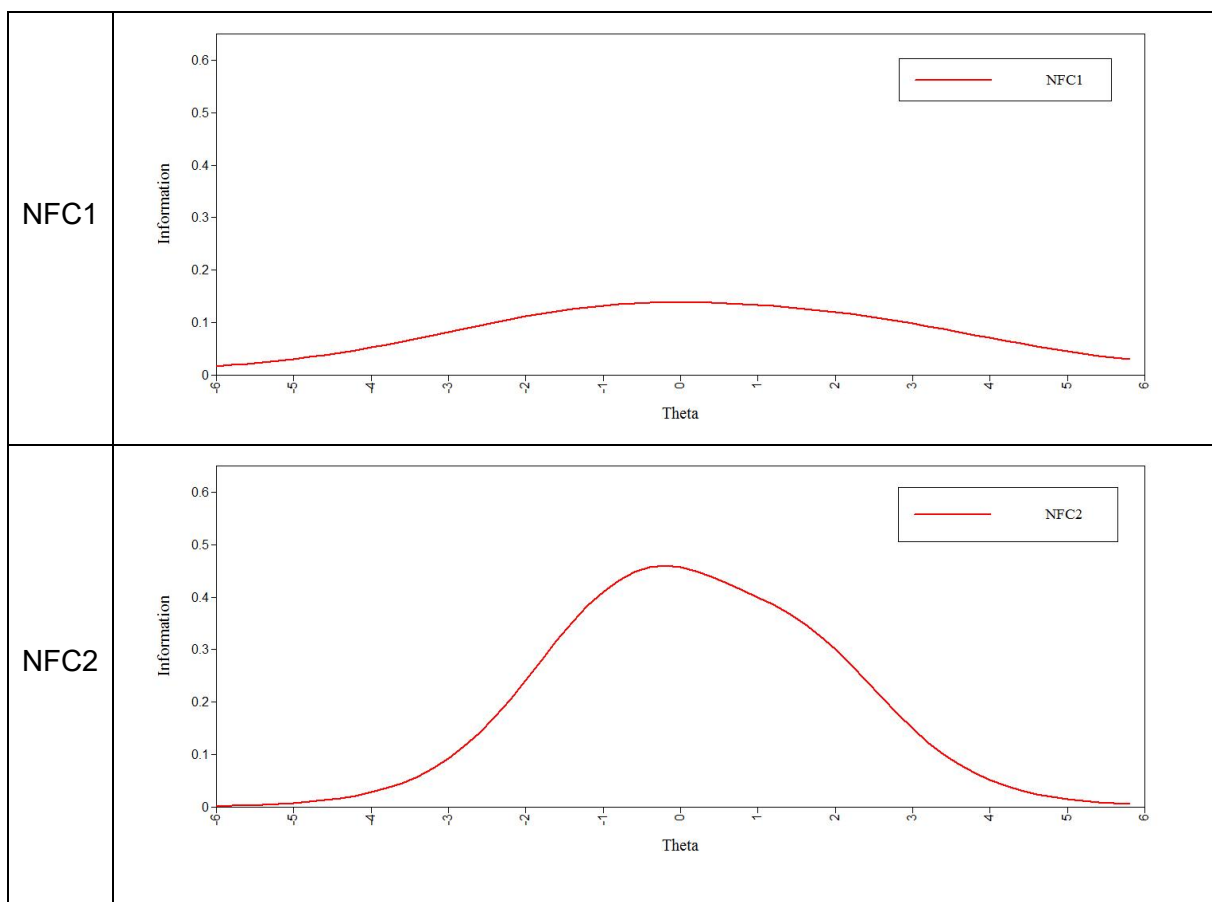
Table 4: NFC scale – Values of discrimination parameters for each item

2.1.2. Step 2 - Item information curve: an indicator of accuracy.

The information curve of an item provides a representation of the accuracy of the latent variable θ . This accuracy is measured by the variability of the item around the central value of θ . The accuracy thus depends on the variance of θ , denoted σ^2 , and is calculated by the formula $1/\sigma^2$. The resulting curve has a Gaussian shape. The accuracy of the item is great when its variance is low. Graphically, when the information curve associated with an item is spread around the central value of θ , the accuracy of the item is great. When the measurement accuracy is insufficient for some items, it is best to remove them.

Rule of interpretation: after reviewing the shape of information curves, items are selected on the basis of a highly constricted curve "pointed" at the central value of the latent trait (θ) because these items show a good accuracy of the information provided.

Illustration with need for cognition scale: the information curve associated with NFC5 has a higher peak for the central value of θ than NFC1 and NFC3 (see Figure 2). The NFC5 item is therefore more precise than NFC1 and NFC3 items on the latent trait. The lack of precision of items NFC1 and NFC3 is an additional argument for a possible elimination of the two items from the need for cognition scale.



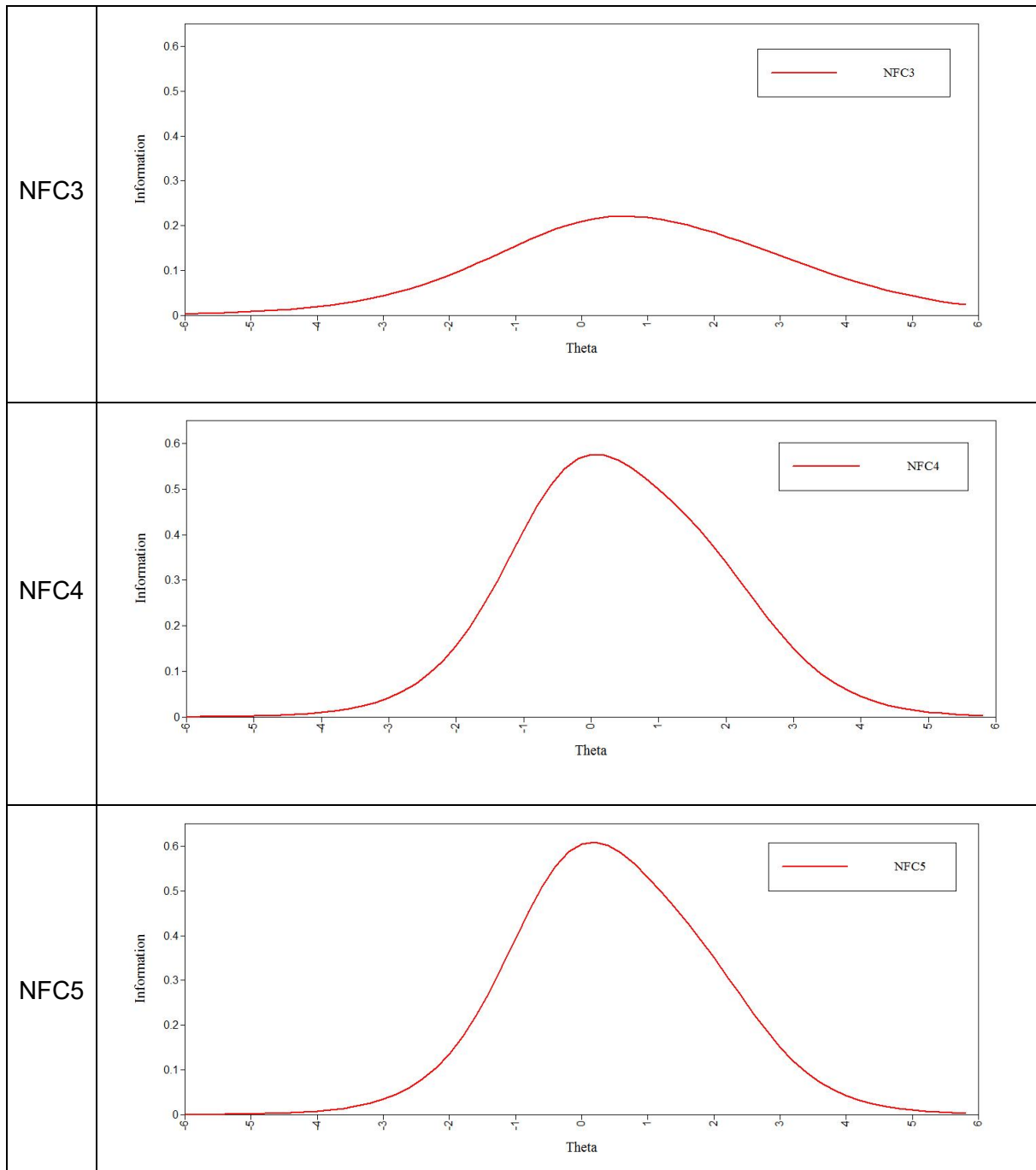


Figure 2: NFC scale – Information curve for each item

Item difficulty and rating points response curves: indicators of difficulty

For multi items scales and rating scales, the difficulty parameter (δ) is measured for each item but also for each rating point of the scale. Each graph is associated with a single item and presents as many curves as the number of rating points of an item. It shows the value θ associated with the item on the horizontal axis and its response probability on the vertical axis. For a rating scale, only the two extreme rating points show a monotonic response curve. Typically, the rating point that has the lowest value among all difficulty has a response curve that decreases monotonically from 1 (for $\theta = -\infty$) to 0 (for $\theta = +\infty$). Conversely, for the rating

point with the highest value, δ has a function which increases monotonically from 0 (for $\theta = -\infty$) to 1 (for $\theta = +\infty$). The other curves are from unimodal distribution correspond to the order of the rating points. Each rating point corresponds to a segment of the latent trait θ for which the response is probable.

The responses curve of each point of an item enables the calculation of the threshold of difficulty associated with the transition from one rating point to another (δ_{ij}). This threshold is the difficulty associated with the transition between two consecutive rating points. Note that an intersection point is the point where two rating points' response curves intersect. Thus there are m intersection points for a $m+1$ rating points, e.g., from "not at all" to "a little" we have one intersection point and another for "a little" to "moderately." Thus there are two intersection points for three rating points scale. The difficulty threshold is measured – in units of θ – at the intersection point. To each difficulty threshold corresponds a value of θ which can be interpreted as the value for which the next point response becomes more probable than the actual one. If the intersection points are well separated from each other, it means that the threshold of difficulty for each point increases, which is desirable (Masters, 1982).

Examination of the response curves for each item rating point allows the identification of rating points considered difficult for respondents because of their lower probability to be selected - spotted lower height than the other. Ideally, the response probability associated with each rating point must be substantially the same, which is characterized by graphic response probabilities $P(\theta)$ relatively equivalent at the intersection curves between consecutive responses.

Rule of interpretation: after examining the response curves of each rating point of an item, retained items are those whose rating point's difficulty will be similar, showing curves of the same height for each point, having a response probability $P(\theta)$ roughly equivalent for all points.

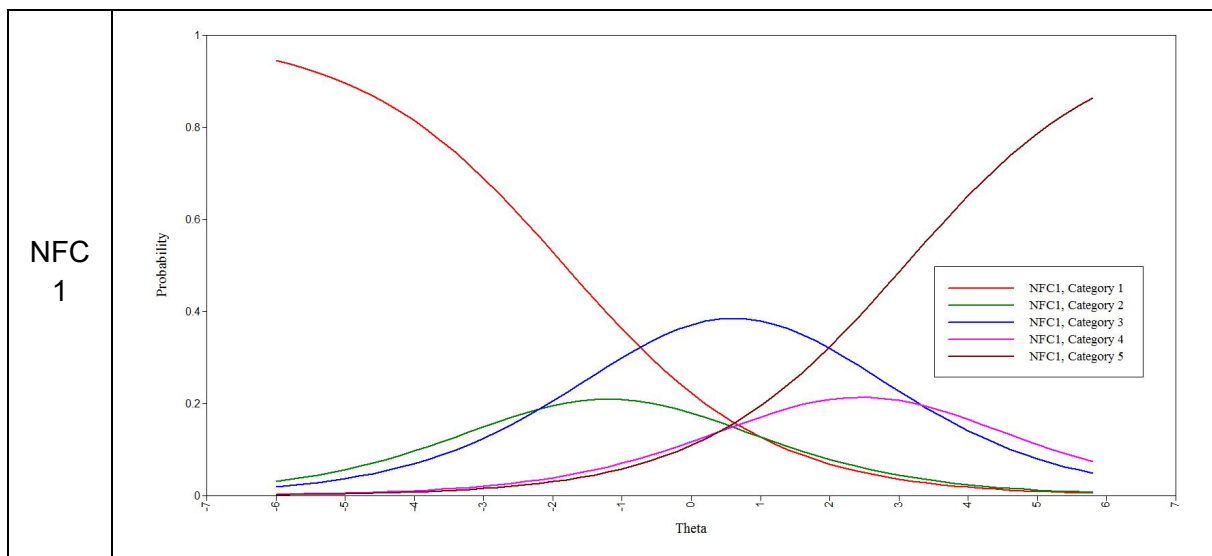
Illustration with need for cognition scale: Figure 3 shows the response curves of the five items. NFC4 presents a rather good response curve showing a satisfying discrimination parameter at each rating point. As NCF 2 et NCF 5 are presenting similar response curves, we discuss NFC5 only. NFC 1 and 3 show very different response curves. Therefore, they are specifically analyzed.

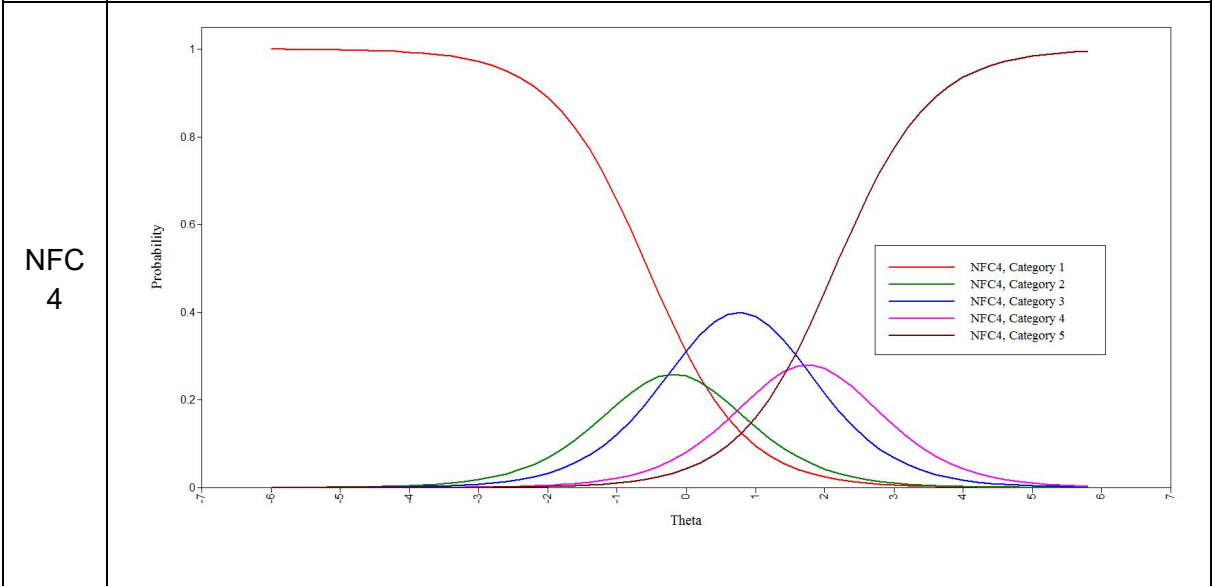
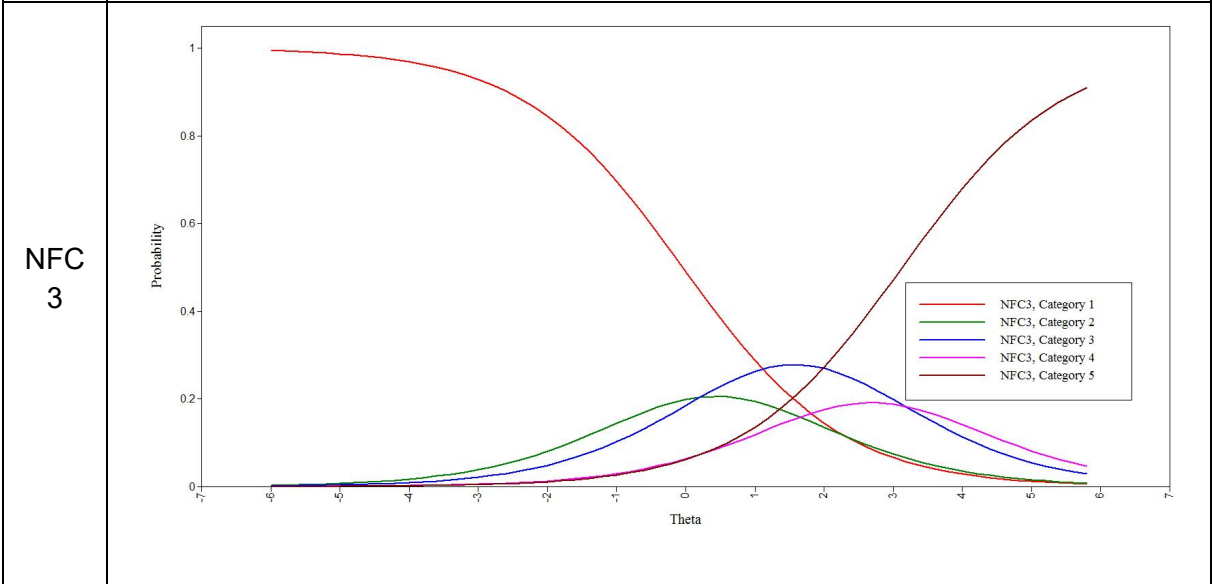
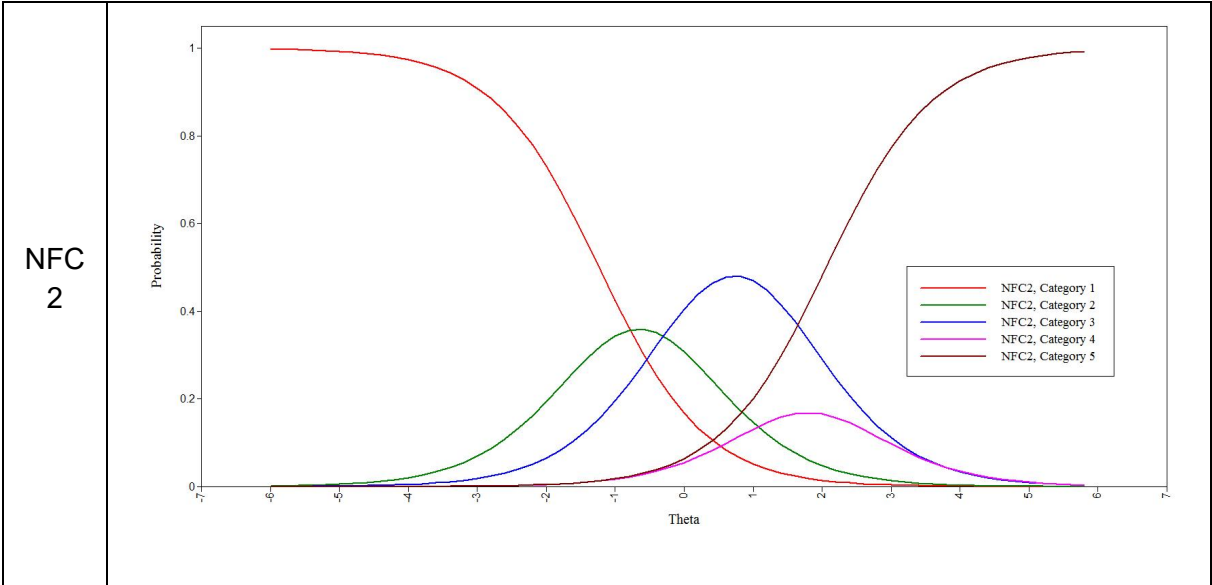
The rating point crossing of NFC 4 curves show a rather regular progressivity on theta (θ). The rating point 3 of the scale shows a slightly higher probability to be selected by the respondents, which is relatively frequent for a neutral point.

The study of the response curves to the 5 rating points of the item NFC5 shows a high and close probability for the rating points 1, 2 and 3. Conversely, going from rating points 3 to 4 has a lower probability than the other intersection points because of its greater response difficulty. To conclude, only rating point 4 of the item NFC5 shows a particular problem in relation to other points that behave normally.

The curves for NFC1 show that rating points 2 and 4 have a lower probability of response due to the greater response difficulty. These rating points have a lower response probability than point 3. Rating point 3 is the central point and has been heavily used by the respondents (more than 40% probability) that did not want or were able to answer the question. Rating point 4 has a low probability of response (about 20 %) due to a greater difficulty. The increase in thresholds of difficulty between consecutive ratings is neither regular nor increasing in θ . Indeed, the intersection points of 2 to 3 and 3 to 4 occurs on more extreme values of θ than the intersection points of 1 to 2 and 4 to 5 indicating that point 3 plays an abnormal role. The item NFC1 needs to be reformulated.

The response curves analysis of item NFC3 rating points show that ratings 2 and 4 have lower responses probabilities (around 20 %) compared to rating 3 (about 30%). Rating 3 also plays an abnormal role. Item 3 appears to be not used as a rating scale. The thresholds of difficulty between levels 2 and 3 and 3 and 4 reached values of θ more extreme than the threshold between levels 1 and 2 or between 4 and 5, indicating that the item NFC3 has a problem of understanding by respondents. However, compared to item NFC1, NFC3 is slightly better in terms of response distribution between the different rating points. The item NFC3 therefore needs to be reformulated.





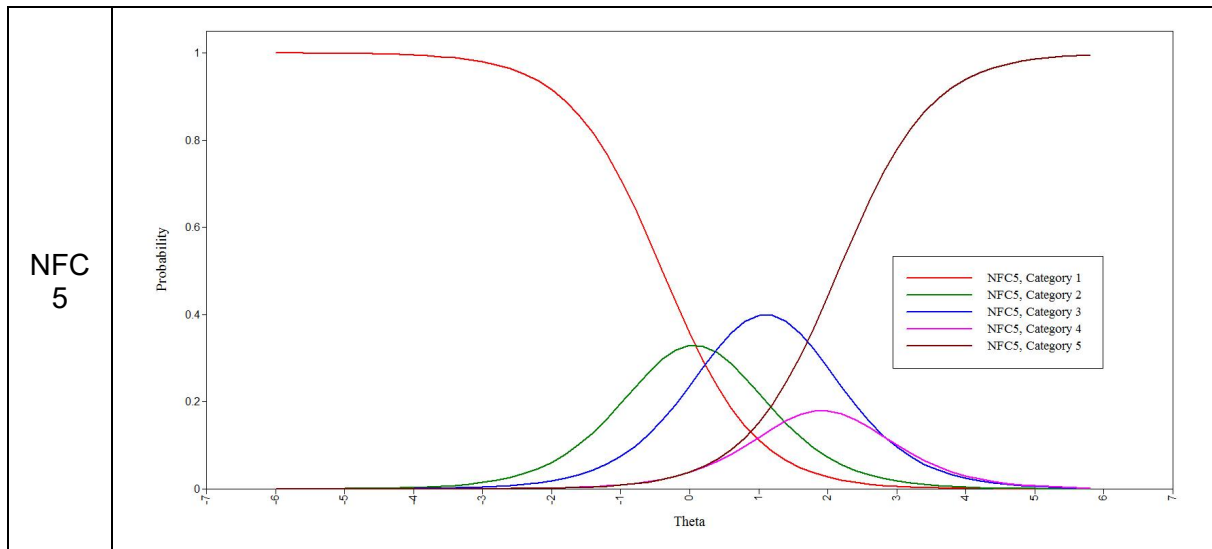


Figure 3: NFC scale – Rating points response curve for NFC items

3. APPLICATION FOR THE NON-NORMAL SCALE OF SATISFACTION

We used the satisfaction scale because it is characterized by the absence of normality already observed by Babin and Griffin (1998). According to the CCT, we performed an exploratory factor analysis, reliability analysis (alpha, rho) and a confirmatory factor analysis using AMOS 18 of a customer satisfaction scale. Mplus software was used to perform the GRM analyses.

The satisfaction scale includes four items. A sample of 151 responses was collected. Each item includes 7 rating points. These range from 1 "not at all satisfied" to 7 "extremely satisfied" for SAT1. SAT2 is measured by a 7-points semantic differential "delighted/terrible." SAT3 and SAT4 are measured using Likert scales from 1 "strongly disagree" to 7 "strongly agree" (see Table 5 for the question labels). SAT2 and SAT3 were reversed, i.e. encoded in reverse order to facilitate the comparison of items.

Items	Responses distribution For each point of the scale									Skewness (test t)	Kurtosis (test t)
	σ	1	2	3	4	5	6	7			
SAT1* - Regarding my relationship with this seller, I am...	5.27	1.28	2	3	13	11	47	55	20	-0.995 (-4.99)	0.943 (2.37)
SAT2** (inverted) - Speaking of satisfaction with my relationship with this seller, I am ...	5.26	1.18	1	4	8	15	54	52	17	-0.926 (-4.64)	1.143 (2.87)
SAT3*** (inverted) - Compared to what I expected, I'm a little disappointed (s) of my relationship with this seller	5.17	1.54	4	8	7	28	27	46	31	-0.813 (-4.08)	0.070 (0.18)
SAT4*** - I am satisfied of my relationships with this seller	5.05	1.43	2	9	10	24	41	43	22	-0.695 (-3.49)	0.000 (0.00)

* 7-point from 1 "not satisfied at all" to 7 "extremely satisfied"
** Semantic differential using 7-point from "delighted" to "terrible"
*** 7-point from 1 "strongly disagree" to 7 "strongly agree"

Table 5: Satisfaction scale – Labels and characteristics of the distribution of items

Normality tests of the satisfaction scale items were performed (see Table 5 for results). The items of this scale are all statistically significant at the 0.05 level for the asymmetry coefficients ($t_{SAT1} = -4.99$, $t_{SAT2} = -4.64$, $t_{SAT3} = -4.08$; $t_{SAT4} = -3.49$). The satisfaction scale is therefore asymmetric and asymptotically shifted to the right. For flattening, SAT1 and SAT2 are sharper than those of a normal distribution at the 0.05 level ($t_{SAT1} = 2.37$, $t_{SAT2} = 2.87$), while SAT3 and SAT4 are normally distributed ($t_{SAT3} = 0.18$, $t_{SAT4} = 0.00$). The coefficient of multinormality (Mardia coefficient) is 9.27 ($t = 8.22$), which confirms the absence of multinormality of the satisfaction scale.

3.1. Results of the classical measurement analysis

3.1.1. Exploratory factor analysis

The results of the exploratory factor analysis of the satisfaction scale are satisfactory with a percentage of the total explained variance of 84.5 % and a quality of representation 0.85 for SAT1, 0.87 for SAT2, 0.84 for SAT3 and 0.82 for SAT4. The Cattell test presents a very strong inflexion point before the second factor, which is a characteristic of a one-dimensional scale. Saturations (or loadings) of the satisfaction scale items are respectively 0.92 for SAT1, 0.93

for SAT2, 0.92 for SAT3 and 0.91 for SAT4. To conclude, these four items are factorable and seem to form a unidimensional scale of satisfaction.

3.1.2. Reliability analysis

The satisfaction scale has good reliability (Cronbach's alpha = 0.93 and rho = 0.79 Jöreskog). If one removes each of the four items successively, the coefficient alpha for three items remains high. Thus, the correlation of the three items with the total score is 0.91 without SAT1, 0.91 without SAT2, 0.92 without SAT3 and 0.92 without SAT4. These high correlations show that the four items are well correlated to the construct of satisfaction (Carmines & Zeller, 1979).

3.1.3. Confirmatory factor analysis

Confirmatory factor analysis is performed on the four items of satisfaction, despite the absence of normality. The results show that the model fits well to the data: $\chi^2/df = 0.80/2$ is not statistically significant with $p = 0.667$, GFI = 0.997, SRMR = 0.048, and RMSEA = 0.000. The standardized Lambdas of each of the four items are all statistically significant and very close to the lambdas obtained by bootstrap: 0.902 (SAT1) and 0.913 (SAT2), 0.883 (SAT3), and 0.805 (SAT4). The score of convergent validity (0.79) is well above the threshold of 0.5, indicating that the scale is valid. These four items are based on the criteria commonly used in marketing and are a very good reflection of satisfaction and the satisfaction scale measures as well. A bootstrap analysis to control non-normality shows very stable lambdas since bootstrap coefficients are very close lambdas standardized 0.903 for SAT1, 0.913 for SAT2, 0.885 for SAT3, and 0.806 for SAT4. In conclusion, the confirmatory factor analysis showed that the four items selected are good indicators of satisfaction.

3.2. Results of the GRM analysis

3.2.1. Step 1 – Item characteristic curve

The characteristic curves of the items show that items SAT1, SAT2, SAT3 and SAT4 present similar results of discrimination (see Figure 4).

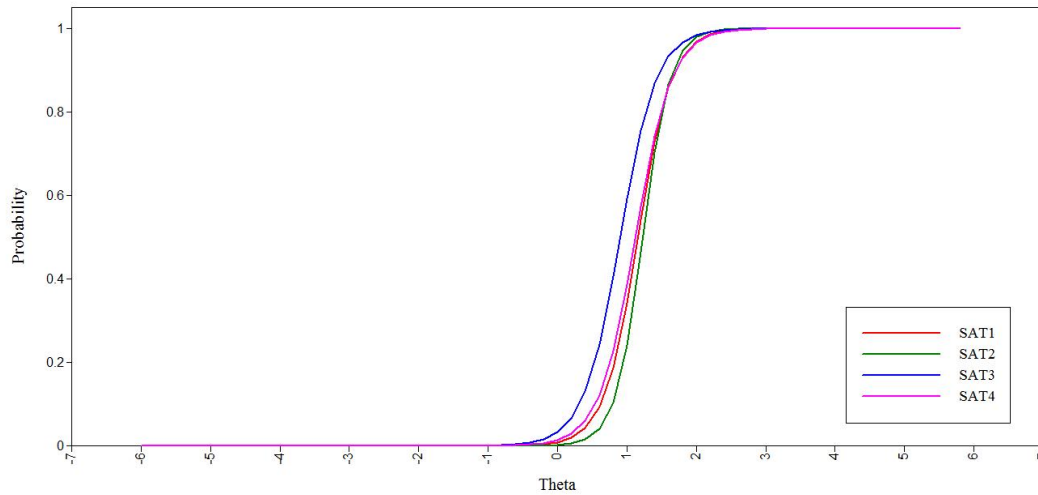


Figure 4: Satisfaction scale – Item characteristic curve for item scales

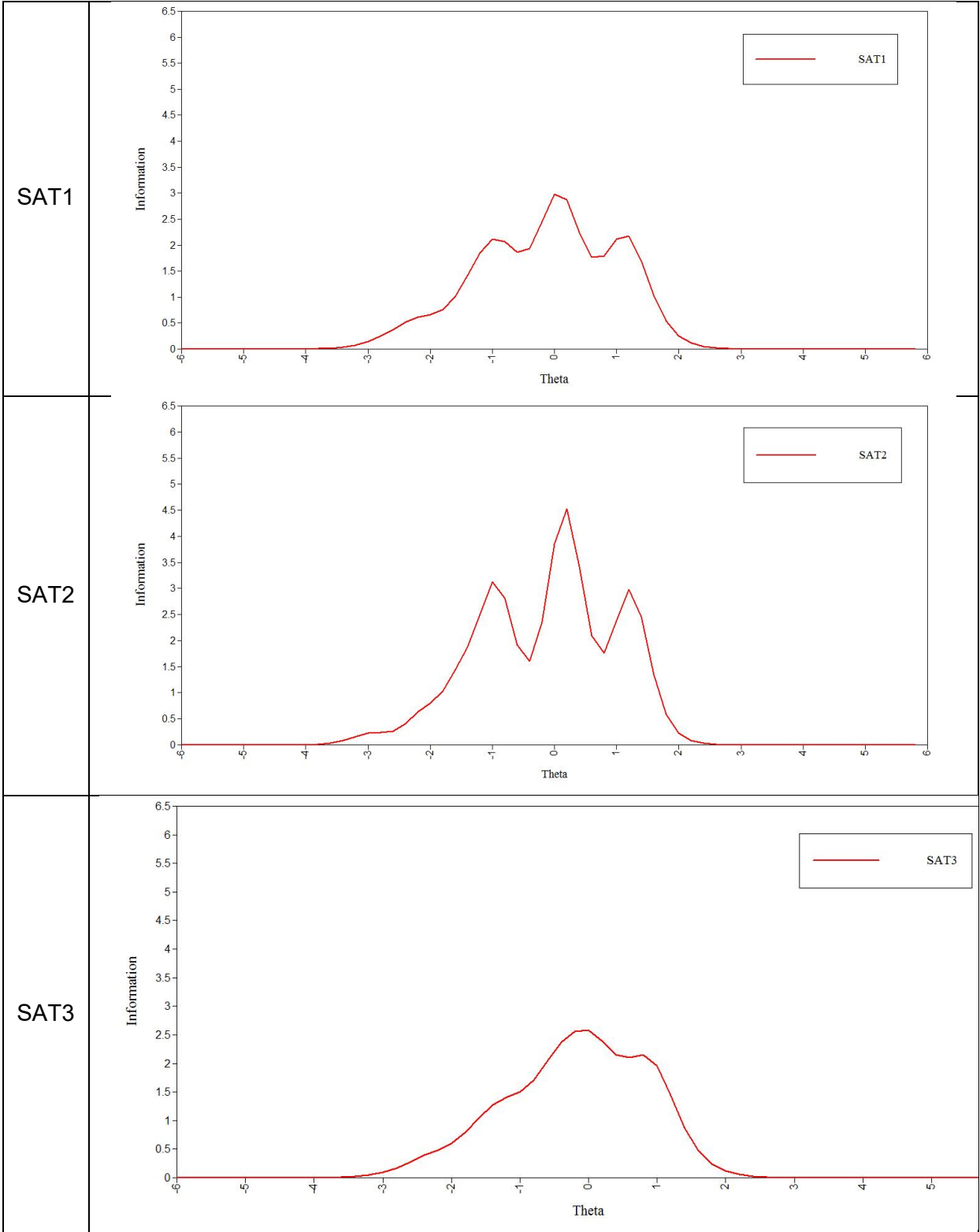
Discriminant α parameters (see Table 6) show that SAT2 is the most discriminant item ($\alpha_{SAT2} = 5.03$). Meanwhile, SAT3 is the least discriminant ($\alpha_{SAT3} = 3.77$). According to Baker's criterion (2001), discriminant parameters of all items are qualified as very strong.

Items	Discrimination parameters (α)	SD
SAT1	4.08	0.66
SAT2	5.03	0.92
SAT3	3.77	0.58
SAT4	3.80	0.62

Table 6: Satisfaction scale – Discriminant parameters of items scale

3.2.2. Step 2 - Item information curve

Information curves of the items show that SAT2 and on a lesser extent SAT1 have a neat peak on θ , meanwhile SAT3 and SAT4 information curves have the flattest curves (see figure 5). It means that SAT2, but also SAT1 have a greater precision than SAT3 and SAT4 on the latent trait.



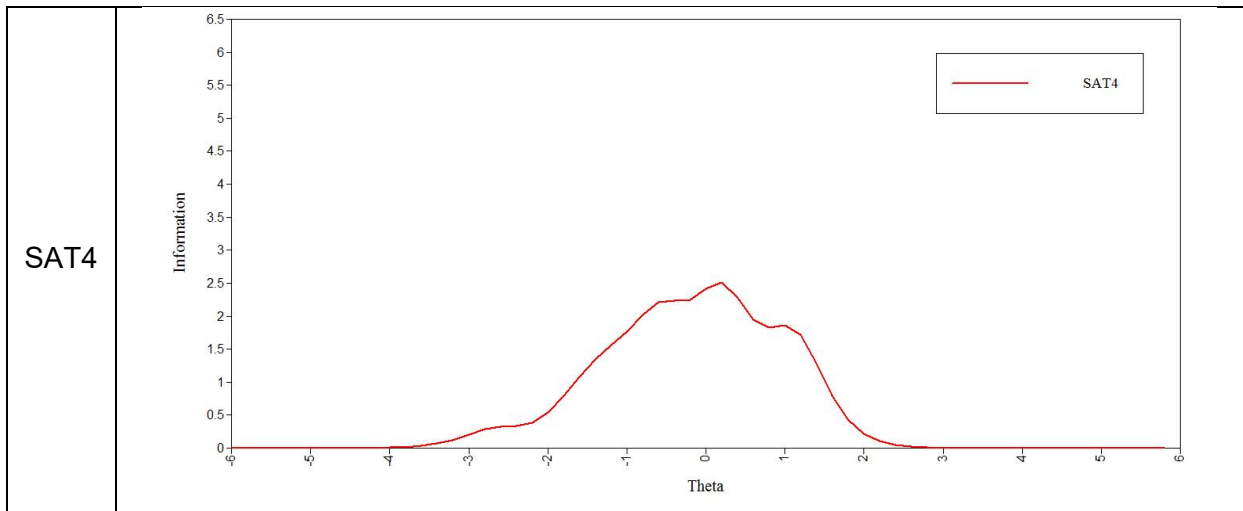
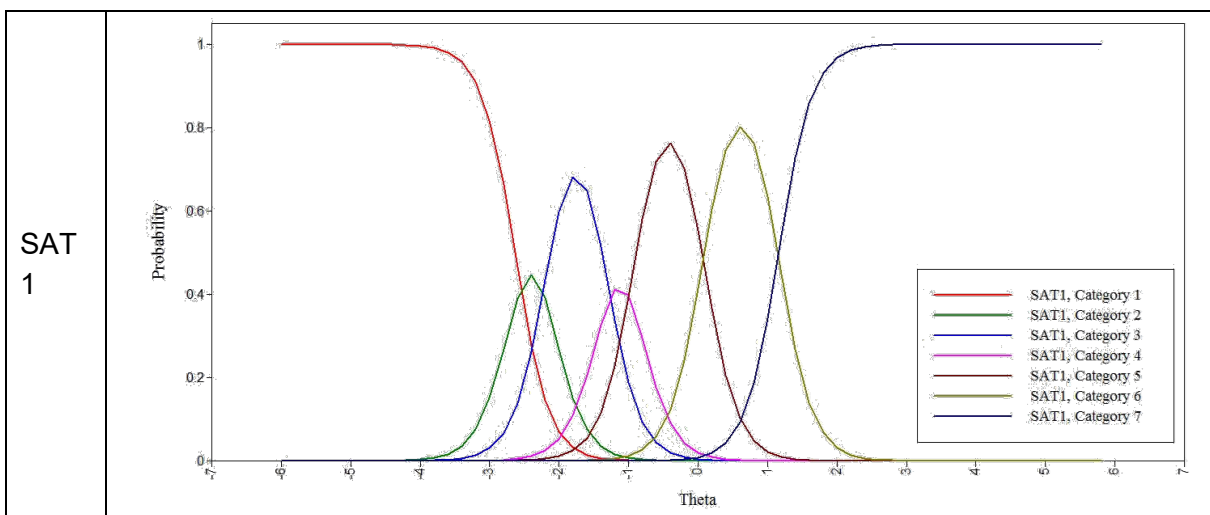


Figure 5: Satisfaction scale (SAT) – Item information curve for each item

3.2.3. Step 3 – Item difficulty and rating point response curves: indicators of difficulty

The different items of the satisfaction scale show very different difficulty parameters. If items SAT1, SAT2 and SAT4 have close difficulty values ($\delta\text{SAT1} = -2.81$, $\delta\text{SAT2} = -2.81$, and $\delta\text{SAT4} = -2.31$), SAT3 has a much less low value ($\delta\text{SAT3} = -1.70$) showing that it is much easier than the other items. In Figure 6, the response curves for each rating point of each item are separated from each other. Intersection points of two rating point response curves show a steady and consistent progress on the latent trait θ . According to Masters' (1982) recommendations, this shows that rating points have relatively equivalent probabilities to be chosen except for rating 3 of SAT3 and SAT4 which have a lower probability. Rating 3 of SAT3 shows also a very small increase on θ . Ratings 2 to 6 of SAT 3 tends to reach a lower probability than the other items, confirming the low difficulty of the item and therefore its unattractive character to measure satisfaction. To conclude, GRM results show that the SAT3 is the least suitable item for measuring satisfaction.



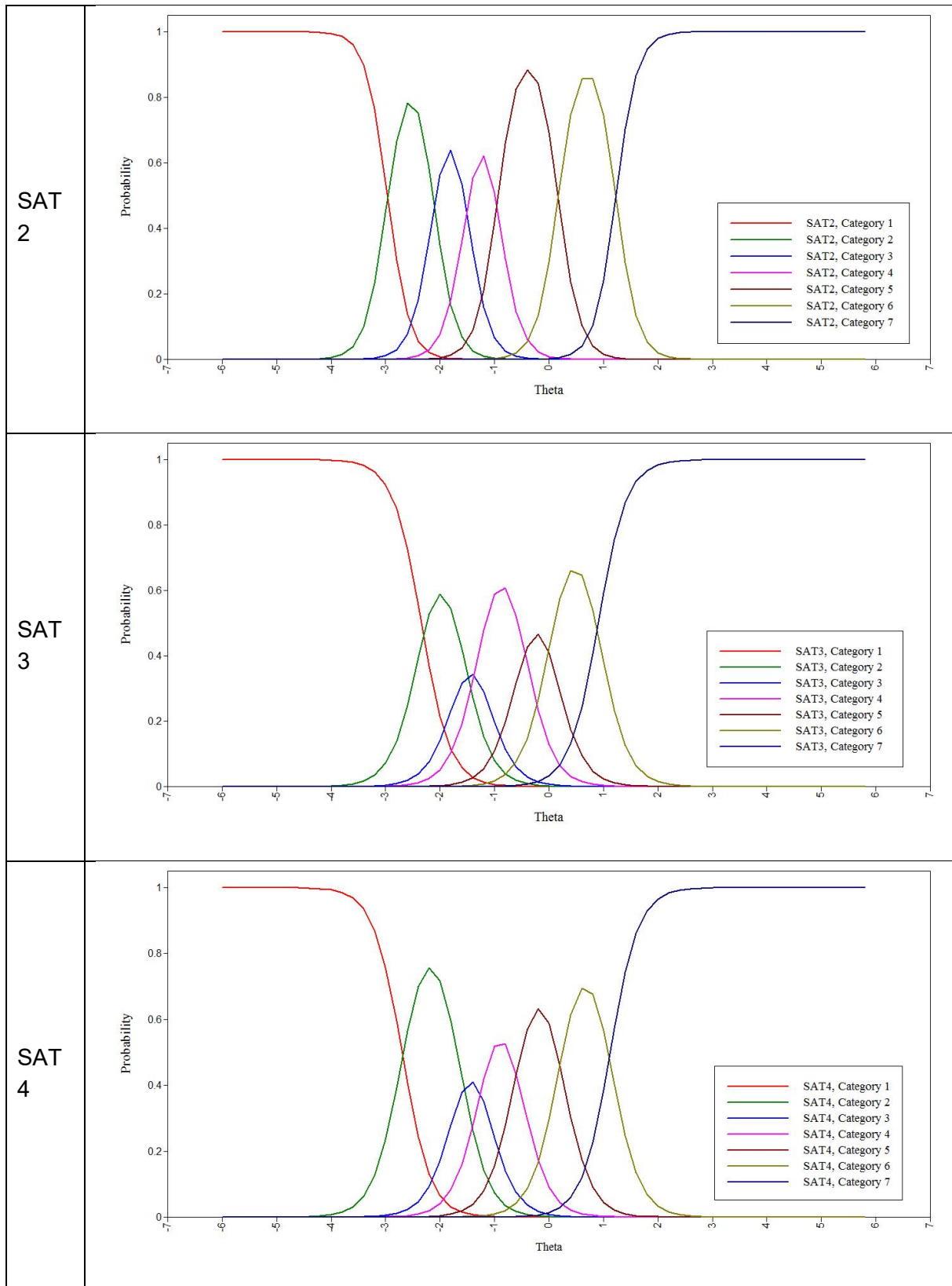


Figure 6: Satisfaction scale – Points response curves for each item

CONCLUSION

GRM analysis allows a more detailed analysis of the items used to measure a latent variable than the structural equation modeling (Reise & Widaman, 1999) because it focuses on each item of the scale of measurement and each rating point used by respondents (Thissen & Steinberg, 1988). Results of the analysis of the NFC scale based on confirmatory factor analysis are not different from those of GRM analysis. Both studies reject NFC1 and NFC3. GRM analysis understands and justifies the removal of an item that is obviously misunderstood by respondents. The confirmatory factor analysis provides the same deletion but needs new data collection. So there is a complementarity between these two methods, and GRM analysis should precede the confirmatory factor analysis on the timeline.

When non-normal distribution occurs on one or more items of a rating scale, the use of GRM analysis is recommended. Results of the satisfaction scale show that the item distributions are characterized by a strong asymmetry. In this case, the use of GRM analysis, which does not assume a linear relationship between the items and the θ latent trait, allows a more rigorous estimate of the scale and its items. GRM analysis shows that only three of the four original items of the satisfaction scale are selected. The item SAT3 is characterized by decreased precision – an information not available using CCT. The item SAT3 focuses on the disappointment felt by the respondent when considering his/her relationship with the salesperson, with an explicit reference to the customer's expectations ("compared to what I expected"). As this item is not a direct measure of customer satisfaction, it is not surprising that the GRM analysis highlights its weaknesses. The satisfaction scale with three items and without SAT3 is thus preferable to having four items.

GRM analysis does not impose any condition on the item distribution and can be performed in the absence of normality. Facing non-normal distribution (skewness or kurtosis), the GRM analysis may complement measurement theory tests to select items (Nunnally, 1994). GRM analysis focuses on each item and its rating points of the measurement scale (Thissen & Steinberg, 1988), which allows a more detailed analysis of the items used to measure a latent trait than that obtained by structural equation modeling (Reise & Widaman 1999). It analyzes in great detail the quality of items of a measurement scale. It refines and thus completes analysis based on classical measurement tests.

GRM analysis can be inserted in the scale construction process. Usually, marketing researchers start from a set of items from which items are selected on the basis of item content (content validity and face validity) and their technical characteristics (construct validity). The GRM analysis provides a set of indicators from which to choose the best items before their use on larger samples. It can be inserted after the analysis of reliability, content validity and face validity and before a confirmatory factor analysis. IRT is used in addition to or as a substitute for the confirmatory factor analysis to ensure that the items and rating points are understood by respondents. It is therefore appropriate to use the GRM analysis to select the items before the confirmatory factor analysis. Overall, GRM analysis promotes a very careful choice of items.

To assess the quality of a scale, GRM parameters are used to distinguish items that are appropriate to measure a given construct in order to avoid item proliferation providing too little additional information. Choosing items that have both similar and high difficulty parameters maximizes the reliability (Lord, 1952). However, it is better to maximize the information that increases the spread of difficulty parameters of the items (Nunnally & Bernstein, 1994), even if the final reliability is lower.

Compared to structural equation modeling, the GRM analysis provides additional information to judge the relevance of the scale point labeling in assessing its difficulty. Often items are rejected for statistical reason only. However, some items may be relevant but very poorly labeled. GRM analysis pinpoints where to improve the wording of the rating points and where a poor understanding of the respondents occurs.

For scales using few items, the researcher usually relies on choosing items on the correlation coefficient between items or when the number of items is equal to 3 only on the value of the Cronbach coefficient to study the reliability of the scale. IRT allows the analysis of the probability of response to a latent trait. It provides new information on the discriminating power of an item to its latent trait, the accuracy of an item and the difficulty associated with the point crossing within each item. It allows the selection of the most representative items of the latent trait, regardless of the normality of the distribution of items.

The use of GRM analysis for the study of one-dimensional rating scales is advocated by Meade and Lautenschlager (2004) due to its complementary information and usefulness for decision-making of managers and researchers. However, GRM analysis is more complex to use than the confirmatory factor analysis for testing multidimensional scales. Note that recent work from Sitjman Paas (2008) is promising and should soon lead to the development of a software testing IRT graded response model for multidimensional scales.

The main obstacle to a widespread application of the GRM analysis is the low number of software and especially their lack of ergonomics. Beyond their scientific interest, their implementation can be somewhat complex and no analytical procedure is currently available in traditional marketing integrated software such as IBM-SPSS or Statistica. Software development is therefore a perspective of major improvement that would encourage the use of this method in marketing.

REFERENCES

- Bacon, L., & Lenk, P. (2008). Breaking the binary model code: IRT models offer flexible marketing measurement. *Marketing Research*, 20(4), 7-10.
- Balasubramanian, S. K., & Kamakura, W. (1989). Measuring consumer attitudes toward the marketplace with tailored interviews. *Journal of Marketing Research*, 26(3), 311-326.
- Baker, F. B. (2001). *The basics of item response theory*, 2nd edition, University of Maryland, College Park, MD., ERIC Clearinghouse on Assessment and Evaluation.

- Bechtel, G. (1985). Generalizing the Rasch model for consumer rating scales. *Marketing Science*, 4(1), 62-75.
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42(1), 116-131.
- Carmines, E. G. & Zeller, R. A. (1979). *Reliability and Validity Assessment*, Thousand Oaks, CA: SAGE Publications.
- Churchill, G. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 16(1), 64-73.
- Cronbach, L. J. (1951). Coefficients Alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-234.
- De Jong, M. J., Steenkamp, J., Fox, J., & Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: a global investigation. *Journal of Marketing Research*, 45(1), 104-115.
- De Jong, M. J., Steenkamp, J., & Veldkamp, B. P. (2009). A model for the construction of country-specific yet internationally comparable short-form marketing scales. *Marketing Science*, 28(4), 674-689.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50.
- Jolibert, A., & Jourdan, P. (2011). *Marketing research : méthodes de recherche et d'études en marketing*, Paris, Dunod.
- Lord, F. (1952). The relationship of the reliability of multiple choice items to the distribution of items difficulties. *Psychometrika*, 18(2), 181-194.
- Mahlhotra, N. K., & Birks, D. F. (2006). *Marketing Research: An Applied Approach*, 2d European edition, Harlow, Prentice Hall.
- Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174.
- Meade, A. W., & Lautenschlager, G. J. (2004). A comparison of item response theory and confirmatory factor analytic methodologies for establishing equivalence/invariance. *Organizational Research Methods*, 7(4), 361-388.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*, Third Edition, New York, Mc Graw Hill.
- Ostini, R., & Nering, M. (2006). *Polytomous item response theory models*, London, Sage.
- Peter, J. P. (1979). Reliability: a review of psychometrics basics and recent marketing practices, *Journal of Marketing Research*, 16(1), 6-17.
- Paas, L., & Sijtsma, K. (2008). Nonparametric item response theory for investigating dimensionality of marketing scales: A SERVQUAL application. *Marketing Letters*, 19(2), 157-170.
- Reise, S., & Widaman, K. (1999). Assessing the fit of measurement models at the individual level: a comparison of item responses theory and covariance structure approaches. *Psychological Methods*, 4(1), 3-21.
- Rossiter, J. R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, 19(4), 305-335.
- Samejima, F. (1996). Evaluation of mathematical models for ordered polychotomous responses. *Behaviormetrika*, 23(1), 17-35.

- Singh, J. (2004). Tackling measurement problems with item response theory: principles, characteristics and assessment with an illustrative example. *Journal of Business Research*, 57(2), 184-208.
- Slijtsma, K., & Hemker, B. (1998). Non parametric polytomous IRT models for invariant item ordering with results for parametric models. *Psychometrika*, 63(2), 183-200.
- Thissen, D., & Steinberg, L. (1988). Data analysis using item response theory. *Psychological Bulletin*, 14(3), 385-395.
- Vrignaud, P. (1996). Les tests au XXIe Siècle : que peut-on attendre des évolutions méthodologiques et technologiques dans la classification et l'évaluation des personnes ? *Pratiques Psychologiques*, 2, 5-28.
- Waller, N., Thompson, J., & Wenk, E. (2000). Using IRT to separate measurement bias to group differences on homogeneous and heterogeneous scales: an illustration with the MMPI. *Psychological Methods*, 5 (1), 125-146.

AUTHORS

Maud Dampérat

Full professor in marketing
Univ Lyon, UJM-Saint-Etienne, COACTIS, EA 4161, F-42023, Saint-Etienne, France

Ping Lei

Assistant professor in marketing
INSEEC School of Business & Economics, F-69007 Lyon, France

Florence Jeannot

Associate professor in marketing
INSEEC School of Business & Economics, CERAG, EA 7521, F-69007 Lyon, France

Alain Jolibert

Emeritus professor in marketing
INSEEC School of Business & Economics, CERAG, EA 7521, F-69007 Lyon, France

L' Apport d'un Logiciel CAQDAS à l'analyse des récits de vie : l'exemple de NVivo

RÉSUMÉ

Les récits de vie sont une méthodologie fortement ancrée dans la sociologie empirique, mais cette méthode offre des perspectives intéressantes pour d'autres disciplines car elle permet de mieux comprendre les interactions entre l'individu, le groupe et la société par la façon dont l'individu se raconte. De plus, les nombreux progrès réalisés par les logiciels d'analyse de données textuelles pour traiter des quantités de données importantes invitent les chercheurs à utiliser les récits de vie. Dans ce chapitre, nous présentons les fondements de la méthode des récits de vie, puis dans une seconde partie nous exposons son traitement avec le logiciel d'analyse de données textuelles NVivo. Notre démarche est autant de répondre aux questions « pourquoi cette méthode ? pourquoi ce logiciel ? » que « comment le faire ? comment les utiliser ? ».

MOTS CLÉS : récit de vie, analyse de données textuelles, principes méthodologiques

INTRODUCTION

Les recherches sur les données qualitatives ont fait de nombreux progrès grâce à l'usage de logiciels d'analyse de données textuelles. Ces logiciels permettent de traiter, à l'instar de leurs cousins quantitatifs, des quantités de données importantes afin d'aider le chercheur à améliorer la compréhension des phénomènes sociaux. Parallèlement, le développement de réseaux sociaux, à dominante personnelle comme professionnelle, encourage les individus à raconter leurs actions, à se raconter. Comment se saisir de la richesse des discours des individus ?

Les manuels en méthodologie de recherche en sciences humaines et sociales expliquent l'art de l'entretien directif, semi-directif ou non directif en explicitant les usages de ces entretiens, la constitution des grilles d'entretien, la conduite de l'entretien et les limites de cette méthodologie. Cependant, la méthode des récits de vie reste peu développée hors de la sociologie empirique alors qu'elle a un potentiel important pour comprendre le comportement des individus dans les autres disciplines des sciences humaines et sociales. De même, la pratique des récits de vie et leur traitement assisté par un logiciel de traitement de données textuelles restent peu abordés dans les ouvrages méthodologiques.

Ce chapitre a vocation à répondre à cet écueil en explicitant les fondements, les usages et les méthodes des récits de vie d'une part, et d'autre part, en présentant leur traitement par un

logiciel d'analyse de données textuelles, NVivo en particulier⁵. Dans une première partie, nous poserons les bases de la méthode des récits de vie, puis dans une seconde partie nous présenterons son traitement avec le logiciel d'analyse de données textuelles NVivo. Notre démarche est autant de répondre aux questions « pourquoi cette méthode ? pourquoi ce logiciel ? » que « comment le faire ? comment les utiliser ? ».

1. LES RÉCITS DE VIE : UN CHOIX DE DONNÉES ET DE POSITIONNEMENT EN RECHERCHE

Le récit de vie est un choix méthodologique comme épistémologique. Dans une démarche inductive, les récits de vie peuvent permettre de révéler des comportements répétitifs, des facteurs déclencheurs et des interactions produisant des réactions similaires afin de découvrir de nouvelles pistes de recherche⁶. Dans une démarche déductive, ils permettent de valider des hypothèses construites à partir des enseignements de la littérature scientifique.

Nous présenterons dans une première partie l'origine de cette méthode, ses spécificités et ses limites. Puis, dans un second temps, nous détaillerons le protocole empirique pour réaliser des récits de vie.

1.1 Révéler le sens de l'action

La méthode des récits de vie émerge dans la sociologie empirique contemporaine à Chicago dans les années 1920-1930 quand les étudiants, encouragés par les enseignants, commencent à travailler sur les récits de vie (Bertaux, 2014). Son origine dans la sociologie s'explique par le fait que « le récit de vie permet (...) d'analyser comment est perçu, senti, vécu et théorisé dans un récit le rapport symbolique qui par le biais du jeu identitaire, relie l'individu au groupe, à la communauté et plus largement à la société » (Chanfrault Duchet, 1987).

Selon Geneviève Pruvost, la méthode biographique repose sur « *un type d'entretien particulier puisqu'il est demandé à quelqu'un de se remémorer sa vie et de raconter son expérience propre. Le dispositif est simple : il n'est pas fondé sur un jeu de questions/réponses à partir d'une grille d'entretien, mais sur l'énoncé d'une consigne initiale qui invite le narrateur à faire le récit de la totalité chronologique de sa vie ou d'une partie, selon l'objectif poursuivi par l'enquêteur. Le sociologue peut faire des relances, poser des questions, mais il doit veiller à ce que l'entretien suive la voie choisie par le narrateur. L'intérêt sociologique du récit de vie réside en effet dans cet ancrage subjectif : il s'agit de saisir les logiques d'action selon le sens*

⁵ Nous avons choisi ce logiciel car il est fortement utilisé dans la communauté scientifique. Plusieurs des points de la méthodologie sont aussi pertinents pour d'autres logiciels similaires, comme Atlas. NVivo fait partie de la famille des logiciels CAQDAS (*Computer Aided Qualitative Data Analysis System*).

⁶ La répétition peut se trouver la fois au sein d'un même entretien quand un.e enquêté.e revient à plusieurs reprises sur un même thème, de la même manière ou non, et entre les différents entretiens. Dans tous les cas, l'identification de la répétition s'inscrit dans une perspective de comparaison.

même que l'acteur confère à sa trajectoire. Loin de singulariser les cas, la méthode du récit de vie permet de situer le réseau dans lequel le narrateur se positionne et d'inscrire les phénomènes sociaux dans un enchaînement de causes et d'effets. Le récit de vie permet de mettre en lumière les processus » (Pruvost, 2010).

Bertaux (2014) recense trois objets qui sont communément appréhendés à partir des récits de vie. Il s'agit d'abord des mondes sociaux qui peuvent par exemple se développer autour d'une activité professionnelle ou d'une activité non-rémunérée. Ainsi, « l'hypothèse centrale de la perspective ethnosociologique est que les logiques qui régissent l'ensemble d'un monde social ou *mésocosme* sont également à l'œuvre dans *chacun* des microcosmes qui le composent : en observant de façon approfondie, un seul, ou mieux quelques-uns de ces derniers, et pour peu qu'on parvienne à en identifier les logiques de situation, les mécanismes générateurs de conduite, les processus de reproduction et de transformation, on devrait pouvoir saisir certaines au moins des logiques sociales du mésocosme lui-même » (Bertaux, 2014, p. 17). Ensuite, les récits de vie permettent d'aborder ce que Bertaux nomme "les catégories de situation" (p. 18), à l'image des mères célibataires. « Cette situation est sociale dans la mesure où elle engendre, pour toutes celles et ceux qui s'y (re)trouvent, plus ou moins les mêmes contraintes, les mêmes tensions, les mêmes logiques (de situation, donc), une pression vers des logiques d'action répondant à la situation et à ses tensions » (Bertaux, 2014) p. 18). On retrouve dans ces deux premiers objets l'idée de répétition déjà évoquée. Enfin, troisième type d'objet pour lesquels les récits de vie sont fréquemment mobilisés, les trajectoires sociales. Selon Bertaux toutefois, il y aurait trop de contingences dans une trajectoire sociale pour qu'elle puisse être appréhendée *via* un récit de vie mais ces récits pourraient en revanche renvoyer à un monde social.

Cette méthode a plusieurs spécificités. D'abord, elle participe d'une démarche auto-réflexive : durant l'échange, le narrateur va construire le récit de sa vie (ou une partie), la personne interrogée/le narrateur, a une capacité à s'auto-analyser et prendre du recul sur son parcours, tout en formulant une cohérence chronologique et de sens à son vécu. Le récit de vie est une manière de se raconter, de se présenter aux autres, elle produit une « identité narrative » (voire une « illusion biographique » pour Bourdieu⁷). Grâce à cette identité narrative, les individus peuvent se faire reconnaître et s'orienter dans leur vie sociale et professionnelle. Ensuite, elle participe à la révélation du sens. Ainsi Bertaux (2014) distingue trois ordres de réalité : i) la réalité historico-empirique, soit celle réellement vécue et agie, ii) le parcours

⁷ Plus précisément Bourdieu souligne que "le fait que « la vie » constitue un tout, un ensemble cohérent et orienté, qui peut et doit être appréhendée comme intention unitaire d'une « intention » subjective et objective, d'un projet » (p. 69) est un "présupposé" (*ibid.*) de cette approche qui ne va pas de soi. Ainsi, "produire une histoire de vie, traiter la vie comme une histoire, c'est-à-dire comme le récit cohérent d'une séquence signifiante et orientée d'événements, c'est peut-être sacrifier à une illusion rhétorique, à une représentation commune de l'existence, que toute une tradition littéraire n'a cessé et ne cesse de renforcer" (*ibid.*, p. 70) et "essayer de comprendre une vie comme une série unique et à soi suffisante d'événements successifs sans autre lien que l'association à un «sujet» dont la constance n'est sans doute que celle d'un nom propre, est à peu près aussi absurde que d'essayer de rendre raison d'un trajet dans le métro sans prendre en compte la structure du réseau, c'est-à-dire la matrice des relations objectives entre les différentes stations » (*ibid.*, p. 71).

biographique, soit la pensée rétrospective du parcours biographique et enfin iii) la réalité discursive qui est produite pendant l'entretien.

Cette méthode nourrit des débats au sujet de l'objectivité, ou à l'inverse de la subjectivité du chercheur. En effet, un récit de vie est co-construit entre le narrateur et la personne à l'écoute, car il est produit dans un contexte de dialogue en interaction avec le chercheur qui irait à l'encontre de l'objectivité des données et de l'analyse. Par exemple pour Demazière et Dubar (dans Bernard, 2014), la parole n'est pas support des représentations déjà formées mais elle les construit, l'entretien qui vise à recueillir les récits de vie est un processus de construction. De même, d'après Patrick Brun, Pierre Bourdieu reconnaît lui aussi l'importance de la relation du chercheur avec l'auteur du récit.

En prolongement de ce processus de construction, dans une perspective interactionniste, le chercheur ne peut prétendre être neutre et s'effacer lors d'un entretien, il est co-constructeur du sens par les questions qu'il pose, ses relances (Bernard, 2014). Par la simple présence du chercheur, et parfois l'orientation de ses questions de relance ou encore sa posture, l'enquêté.e réagit de manière singulière, risquant d'omettre ou d'exagérer certains points. Cette relation peut donc impacter le récit dans son contenu et dans sa forme. Cependant, c'est bien d'abord « le narrateur qui connaît le dénouement de son récit [qui] sélectionne les éléments qu'il juge pertinents pour faire comprendre le sens de celui-ci » (Bernard, 2014, p. 2). A travers ce processus de construction du récit deux types d'éléments sont triés ; ainsi, « les récits témoignent de l'aptitude des acteurs sociaux à recomposer le passé (le temps) et à faire des choix mnésiques (la mémoire) » (Bernard, 2014, p. 2).

Finalement, l'enquêté.e pourrait être renommé.e "le narrateur" ou "la narratrice" dès que l'on considère ses paroles non plus comme des données mais bien comme un récit construit, d'où la qualification de l'entretien visant à recueillir les récits de vie comme "entretien narratif" par Bertaux (2014, p. 10). Ainsi, « Ricoeur analyse le récit comme une « mise en intrigue » : en faisant le récit de sa vie, un individu reconfigure son expérience temporelle, à l'origine confuse, autour d'une intrigue, qui transforme les événements vécus en une histoire. En outre, il est conduit à sélectionner parmi les nombreux événements ceux qui lui semblent « significatifs » » (Giraud et al., 2014, p. 4). Cette construction renvoie ainsi à la mimesis II qui se définit comme "le moment de la mise en intrigue proprement dite par l'auteur du texte. La dimension épisodique du récit tire le temps narratif du côté de la représentation linéaire : les épisodes se suivent l'un l'autre en accord avec l'ordre irréversible du temps commun aux événements physiques et humains. La dimension configurante, en retour, présente des traits temporels inverses de ceux de la dimension épisodique : « l'arrangement configurant transforme la succession des événements en une totalité signifiante qui est le corrélat d'assembler les événements et fait que l'histoire se laisse suivre. Grâce à cet acte réflexif, l'histoire peut être traduite en une « pensée » ». De plus, « la configuration de l'intrigue impose à la suite indéfinie des incidents « le sens d'un point final » [...]. En lisant la fin dans le commencement et le commencement dans la fin, nous apprenons ainsi à lire le temps lui-même à rebours, comme la récapitulation des conditions initiales d'un cours d'action dans ses conséquences terminales » (Ricoeur, 1983, p. 130-31). Ainsi, la mimesis II transfigure l'amont du texte en aval par son pouvoir de configuration. Elle renvoie à la temporalité de la narration" (Carcassonne, 1984, p. 54). Plus directement dans les termes de Ricoeur, "la mise en intrigue a été définie, au plan le plus formel, comme un dynamisme intégrateur qui tire une histoire une et complète d'un divers

d'incidents, autant dire transforme ce divers en une histoire une et complète » (Ricoeur, 1984, p. 18).

Bien qu'en admettant son caractère construit, le récit de vie comme procédé d'enquête pose question.

Une des limites mises en avant par les sceptiques de la méthode biographique est caractérisée par le risque de subjectivité du chercheur dans le traitement du matériau. Les chercheurs tendent à être le plus objectif possible, mais tous impliquent leur subjectivité dans la manière de questionner puis d'analyser et interpréter le récit d'un sujet. Cependant, cette subjectivité pourrait être exacerbée suite à un entretien qualitatif, puisqu'il s'agit tout d'abord d'une rencontre humaine. Le risque majeur serait enfin que le chercheur pensant détenir la vérité, interprète cet entretien selon son jugement qu'il pense être le plus vrai, et risque de passer à côté de la « vérité » du terrain. C'est d'autant plus le cas dans une démarche déductive où certaines hypothèses sont déjà explicitement posées et peuvent ainsi réduire l'interprétation des éléments de sens identifiés à ceux qui sont déjà identifiés dans la littérature scientifique ou plus radicalement conduire à ignorer certains éléments de sens car ils ne rentrent pas dans le cadre théorique préalablement énoncé.

Finalement, « le matériau est éminemment qualitatif. Cependant la frontière avec les méthodes quantitatives n'est pas toujours aussi nette » déclare Blandine Veit (2004). En effet, il est préférable avec une telle démarche d'emprunter à plusieurs méthodes pour analyser le matériau spécifique qui nous est donné à voir avec les récits de vie. Il n'est donc pas exclu de recourir aux méthodes quantitatives, à l'analyse de discours, et à d'autres disciplines comme la littérature, la linguistique et la philosophie. Il semble aussi pertinent d'établir des calendriers, ainsi que de comparer les récits du point de vue des logiques individuelles ou collectives (Veit, 2004). De même, la mobilisation de plusieurs types de sources ou supports peut permettre de se rapprocher avec plus de justesse de la réalité empirique quand c'est à cette dernière que l'on s'intéresse plus qu'à la perception qu'en a l'enquêté.e., ou d'approfondir les différentes manières selon lesquelles un individu peut se raconter selon le contexte d'énonciation. Ainsi, l'analyse des profils sur les réseaux sociaux peut par exemple compléter le récit que l'individu fait de lui-même face à l'enquêteur.trice par une autre forme de récit destiné à un autre type de public.

1.2. L'art du récit de vie ou précis méthodologique pour réaliser sa collecte de données

Après cette brève revue du positionnement de recherche qu'impliquent les récits de vie, l'objectif de cette seconde sous-partie est de présenter plus concrètement quelques astuces dans la mobilisation de cette méthode. D'abord, la collecte des données correspond le plus souvent à une situation d'entretien. Face à l'enquêté.e, on peut privilégier les formules orales : rien ne sert de parler "comme un livre", cela risquerait plutôt de gêner voire déstabiliser l'enquêté.e, selon le niveau de langue qu'il ou elle a l'habitude d'employer. Même si l'entretien des récits de vie est une situation d'interaction particulière par ses modalités, elle reste un échange oral entre deux "humains". On peut par exemple préférer les formules du type "est-ce que vous pourriez m'en dire plus sur...?" par rapport aux formules du type "pourriez-vous m'en dire plus sur...?" ou pire, parce qu'encore plus soutenu, "vous serait-il possible de m'en

dire plus sur...?”. Si le plus neutre est peut-être de s’adapter au niveau de langue de l’enquêté.e afin que ce point ne constitue pas une gêne à l’entretien, en l’absence de connaissance de cette caractéristique de l’enquêté.e on peut utiliser un langage oral courant.

Dans la mesure où l’entretien telle qu’il est le plus souvent pratiqué en contexte de recherche est une situation de déséquilibre entre un.e enquêteur/trice qui est censé.e intervenir peu et se contenter de poser des questions et un.e enquêté.e qui est invité.e à parler plus longuement, voire très longuement, pour répondre aux questions, il est préférable d’éviter tout ce qui pourrait ajouter au malaise dû à la particularité de la situation. La difficulté est d’autant plus importante dans le cadre de l’approche biographique où l’enquêté.e n’est pas nécessairement interrogé.e en tant qu’expert.e sur un sujet qu’il ou elle maîtriserait ou sur un contexte dans lequel il ou elle évolue mais bien plus directement sur lui-même ou elle-même⁸. Dans le cas précis des entretiens visant à collecter les récits de vie, la formulation « racontez-moi... » a cela de particulièrement pertinent qu’elle invite au récit (Bernard, 2014).

En réponse aux limites de l’entretien biographique évoquées plus haut, il est conseillé d’utiliser des formules aussi neutres que possibles pour relancer l’enquêté.e. mais sans oublier, encore une fois, que les deux interlocuteurs sont humains donc soumis à un certain nombre d’aléas. Selon sa concentration, sa disponibilité mentale ou le stade de sa recherche l’enquêteur/trice sera par exemple plus ou moins performant.e dans son entretien - étant entendu que le terme de performance appliqué à un entretien effectué dans le cadre d’une recherche renvoie à la capacité de l’enquêteur/trice à rester neutre tout en faisant parler l’enquêté.e. De même, certain.e.s enquêté.e.s sont plus ou moins enclin.e.s aux digressions, aux réponses courtes ou évasives, ou peuvent être plus ou moins disponibles ou agacé.e.s par la situation de l’entretien...tout en l’ayant préalablement acceptée⁹.

Nous proposons ici quelques stratégies de relance qui semblent pouvoir s’appliquer quelle que soit la thématique de l’entretien, tout en soulignant qu’une bonne stratégie est une stratégie qui fonctionne dans le cadre d’une recherche particulière¹⁰. D’abord, la stratégie qui se présente probablement comme la plus évidente *a priori* pour un.e enquêteur/trice qui débiterait sa recherche consiste à prévoir à l’avance les phrases de relance et les noter sur le guide d’entretien. Cela permet d’anticiper à tête reposée les éventuels biais de formulation et de pouvoir se reposer sur le travail préalable au moment de l’interaction qui peut être stressante. Ensuite, les formules de demande de précision vague du type “c’est-à-dire?” ou

⁸ Cependant, pour Bertaux (2014), l’entretien biographique, dans une perspective ethnosociologique, vise bien, au contraire, à recueillir des données non pas sur l’individu et sa subjectivité mais à travers lui sur le contexte dans lequel il évolue, comme une sorte de témoignage d’un monde social. De même pour Giraud et al. « il s’agit pour le sociologue de retracer l’histoire individuelle de l’acteur afin de mettre au jour les cadres socialisateurs différenciés qu’il a traversés et les marques, sous forme de manières de penser ou d’agir, laissées en lui par leur fréquentation » (2014, p. 3).

⁹ On suppose en effet que l’enquêté.e a bien été informé de la nature de l’enquête et de son contexte et a ainsi pu l’accepter en connaissance de cause.

¹⁰ Cette remarque renvoie non seulement au fait que certains objets d’enquête peuvent appeler certaines relances particulières mais aussi au fait que l’enquêteur/trice en tant qu’individu peut être plus à l’aise avec certaines stratégies que d’autres, de même que l’enquêté.e.

“est-ce que vous pouvez m’en dire plus là-dessus?” sont également très faciles à mobiliser. Enfin, la répétition des paroles de l’enquêté.e demande peut-être plus de concentration et d’habileté : soit on rebondit sur les dernières paroles prononcées soit, si l’on prend des notes en même temps que le narrateur ou la narratrice déroule son récit, on peut revenir sur un élément évoqué plus tôt. Il est préférable toutefois de bien porter attention à reprendre les termes employés par l’enquêté.e, afin d’éviter de déjà plaquer sa propre interprétation et parfois même directement son propre cadre d’analyse sur les paroles de l’enquêté.e qui risquerait ensuite s’en emparer alors que les termes de l’enquêteur/trice ne font pas partie de ses références spontanées. Concrètement, on peut répéter telles quelles les paroles de l’enquêté.e (voir encadré 1).

Enquêté.e : et donc c’est là que je me suis dit qu’il fallait que je change de métier.
Enquêteur/trice : vous vous êtes dits qu’il fallait que vous changiez de métier...
On peut aussi y ajouter une des questions de relance précédentes :
Enquêté.e : et donc c’est là que je me suis dit qu’il fallait que je change de métier.
Enquêteur/trice : vous vous êtes dit qu’il fallait que vous changiez de métier, est-ce que vous pouvez m’en dire plus (là-dessus) ?

Parfois un simple “ah?” ou “ah bon?” ou même “hum?” suffisent à relancer l’enquêté.e, sans oublier aussi que toute situation de communication implique un langage non-verbal: une expression d’interrogation sur le visage peut ainsi appeler l’enquêté.e à poursuivre son propos, à détailler plus la situation qu’il ou elle évoque, à apporter des précisions sur la thématique abordée. De même la prise de note face à l’enquêté.e au moment de l’entretien peut influencer sur son débit de parole en l’encourageant lorsque l’enquêteur/trice prend des notes ou au contraire en le restreignant lorsqu’il n’en prend pas¹¹.

Les temps de silence ont également un rôle important : ils permettent d’une part à l’enquêté.e de réfléchir à ce qu’il va ou vient de répondre et d’autre part ils l’incitent à parler : étant le plus souvent inhabituels dans une conversation spontanée, l’enquêté.e tendra à vouloir les combler en parlant. L’enquêteur/trice doit donc se montrer habile dans l’utilisation des silences et savoir les manier sans susciter le malaise de l’enquêté.e ce qui risquerait d’impacter ses paroles. Concrètement trois à quatre secondes de silence sont suffisamment inhabituelles sans devenir pour autant gênantes. Ce temps de silence montre également que l’enquêteur/trice est à l’écoute de l’enquêté.e et ne cherche pas à l’interrompre ou à passer à la rubrique suivante de manière mécanique, comme cela peut parfois être le cas lors des premiers entretiens menés dans le cadre d’une nouvelle ou première recherche. En effet, il faut le plus souvent un temps de rodage du guide d’entretien et de familiarisation avec les réactions qu’il suscite pour éventuellement l’adapter et tout du moins l’utiliser à bon escient. En résumé, la “règle d’or de l’entretien narratif” (Bertaux, 2014, p. 49) consiste à “encourager le sujet à parler, par de simples approbations et relances, en l’interrompant le moins souvent possible” (*ibid.*).

¹¹ Cela dit, il n’y a pas de règle quant à la réaction des enquêté.e.s face à la prise de notes et certain.e.s pourront par exemple tout autant ralentir leur débit si l’enquêteur/trice note rapidement dans le but de l’aider dans sa prise de note, tandis que d’autre pourront y être complètement indifférent.e.s.

2. TRAITER ET ANALYSER DES RÉCITS DE VIE AVEC UN LOGICIEL D'ANALYSE DES DONNÉES TEXTUELLES

A l'instar des logiciels de traitement des grands nombres, il existe aujourd'hui des logiciels de traitement et d'analyse des données qualitatives. Ces logiciels permettent d'améliorer le processus de traitement des données (Hutchison et al, 2010) en augmentant l'efficacité du traitement. Cette aide logicielle ne se substitue par au travail d'interprétation du chercheur, mais améliore l'organisation et le traitement systématique des données. Nous avons choisi de présenter le logiciel NVivo, très utilisé dans la communauté scientifique, dans le cas particulier des récits de vie. En effet, ce logiciel pour avoir d'autres usages pour le chercheur : traitement des images, mais aussi analyse d'une revue de littérature. Dans un premier temps, nous rappellerons les fondements et les usages de l'analyse textuelle, puis dans un second temps, nous présenterons l'utilisation du logiciel dans le traitement de récits de vie.

2.1 Analyse textuelle avec un logiciel CAQDAS

Avant de commencer le traitement, le chercheur doit se poser plusieurs questions. La première concerne la prise en compte de la temporalité. On peut rencontrer plusieurs types de références au temps mobilisées par les enquêtés : la temporalité en tant que telle, marquée par des formules telle qu'« à l'époque », la mémoire avec les « je me souviens » et la rétroaction qui consiste à mettre en sens au moment de l'entretien un élément passé, avec des formulations comme « j'en parle maintenant, je ne le voyais pas avant » (Bernard, 2014). La seconde est liée aux données : dans une approche structuralisme, le chercheur considère l'ensemble de textes comme homogène, comme un sac de mots qu'il s'agit d'analyser, à l'inverse, dans l'approche sociolinguistique interactionniste la retranscription d'entretien est considérée comme hétérogène, l'entretien ne produisant pas des données brutes mais des productions narratives avec un début, un milieu, une fin (Bernard, 2014). Rien n'interdit cependant de tester les deux approches sur un même ensemble de texte pour les comparer. Une fois ces décisions épistémologiques prises, le travail de nettoyage et de traitement des données peut commencer.

L'analyse textuelle consiste dans un premier temps à recenser, classer et quantifier les traits d'un corpus. Ce traitement est d'abord descriptif en recensant et identifiant les thèmes abordés dans le corpus, et en produisant un décompte fréquentiel du vocabulaire pour aboutir à une analyse statistique. Ensuite, il s'agit de faire apparaître ces liens en croisant les éléments entre eux, et d'étudier aussi les concomitances thématiques comme par exemple : « il parle de son métier il aborde toujours aussitôt la question de l'argent ».

Le formulaire d'analyse est l'instrument avec lequel on interroge le corpus. Il va du répertoriage de quelques thèmes privilégiés dans l'analyse thématique classique, à l'établissement d'un véritable questionnaire avec ses groupes de questions, ses échelles, ses indices numériques. Les chercheurs doivent utiliser une grille à au moins trois degrés. Par exemple, le thème étant "les relations familiales dans votre enfance" : i) les membres de la famille (père, mère, le sujet,

un frère), ii) les types de relation (d'autorité, d'information, agressives, d'aide, affectives, corporelles), iii) les interrelations (père-mère, père-sujet, père-frère etc.)¹².

Pour réaliser le traitement, le logiciel NVivo permet de coder des textes, retranscriptions d'entretien ou autres, c'est-à-dire d'attribuer à des extraits dont on choisit la longueur un ou plusieurs mot(s)-clé(s) appelés "noeuds" sur le logiciel. Mais ce procédé en lui-même pose plusieurs questions. D'abord, quelle échelle choisir, c'est-à-dire quelle taille d'extrait est la plus pertinente pour l'analyse ? Glaser et Strauss (dans Point et Voynnet-Fourboul, 2006) suggèrent non pas tant de choisir *a priori* une échelle mais plutôt de découper le texte en fonction de la survenue d'un "incident critique" (p. 64) entendu comme "une idée, ou un ensemble d'idées isolables par rapport au reste des données qualitatives et qui représente une certaine cohésion" (*ibid.*). Plus qu'un outil mécanique qui s'appliquerait sans réflexion, le découpage d'incidents critiques repose en partie sur l'intuition ; une précaution à prendre étant de conserver dans chaque extrait des éléments de contexte afin d'éviter de dénaturer les propos de l'enquêté.e au moment de l'analyse proprement dite¹³. Chaque extrait peut être codé plusieurs fois, c'est-à-dire se voir attribuer plusieurs noeuds : l'intérêt est alors de pouvoir faire ressortir d'éventuelles cooccurrences entre les différents codes. Les différents codages peuvent s'empiler sans se superposer en termes de bornes de l'extrait : un extrait A peut se terminer par une phrase qui appartient aussi à l'extrait B et le début de l'extrait B n'a pas à remonter au début de l'extrait A. Sur la figure 1 par exemple, on peut observer sur la droite de la capture écran les bandes de différentes couleurs représentant les passages codés.

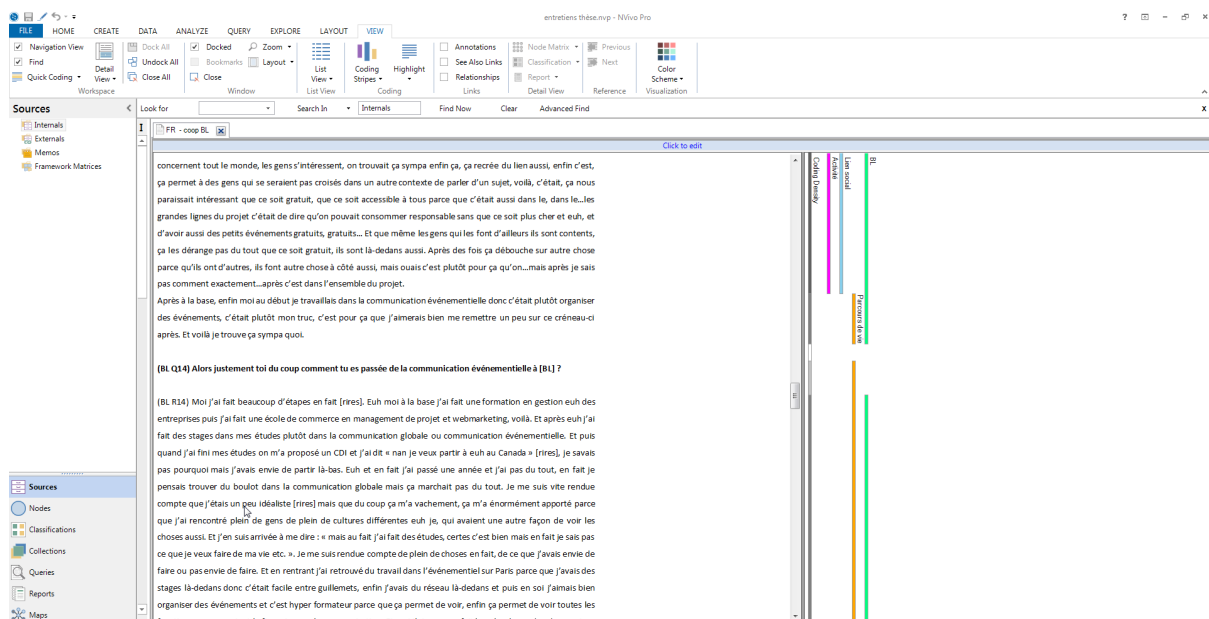


Figure 1 : Superposition de codage sur NVivo

¹² Source : <http://www.europsy.org/marc-alain/analysecontenu.html> et <http://www.suristat.org/article312.html>

¹³ On distingue en effet bien le codage, étape d'attribution de catégorie à un code, de l'analyse qui renvoie plus globalement à la fois au codage et à l'interprétation ou l'utilisation qui en est ensuite faite par le chercheur ou la chercheuse.

Ensuite, quelle démarche utiliser pour mettre en place ces codes ? On retrouve ici en jeu les deux grandes démarches de recherche : hypothético-déductive et inductive. En accord avec notamment Point et Voynnet, nous conseillons d'adopter une posture à mi-chemin entre les deux, ce qu'ils nomment le "codage à visée théorique" (2006, p. 64) et qui consiste à "coupler ces deux approches, afin de procéder de manière itérative. Il s'agit alors de favoriser des « allers et retours » entre les données recueillies et les interprétations successives du chercheur, ce que nous assimilons à une démarche d'abduction" (*ibid.*). Il est également possible, au fur et à mesure de l'avancement du codage de procéder à des ajustements entre les nœuds en créant par exemple des sous-codes mais aussi en précisant ou amendant la description d'un code, ce que le logiciel permet de faire puisqu'au-delà de l'intitulé du code un paragraphe de description lui est consacré. Par exemple, on peut se rendre compte qu'un nœud est trop vaste et renvoie à une diversité de situations qui sont parfois contradictoires entre elles et rendent finalement le nœud non-pertinent en termes d'analyse, d'où la nécessité de le subdiviser. Ainsi le nœud "travail" décrit comme "tous les éléments sur le rapport au travail" peut être décomposé entre "satisfaction au travail" et "insatisfaction", nœud qui peut à son tour donner lieu aux nœuds "insatisfaction des conditions matérielles du travail" et « absence de réalisation de soi ». Stéphanie Abrial (2019) conseille de s'appuyer sur plusieurs stratégies de codage renvoyant aux codes thématiques, transversaux, contextuels/émotionnels et séquentiels. Les premiers se rapportent aux éléments en lien avec la recherche ; les deuxièmes aux éléments *a priori* sans rapport avec la recherche mais que l'on est susceptible de retrouver dans tout entretien, par exemple les questions autour de l'argent ; les troisièmes peuvent décrire tant les conditions de réalisation de l'entretien, par exemple le lieu où il se déroule, que les réactions de l'enquêté.e, par exemple ses rires ou ses silences ; les quatrièmes, enfin, font apparaître la structure de l'entretien. Quand le guide d'entretien est rigoureusement suivi, cette dernière stratégie mène à identifier des codes qui lui font directement écho.

Enfin, une limite, ou en tout cas un rappel, qui mérite d'être souligné est que le logiciel ne code que ce qu'on lui demande de coder et de la manière dont on lui demande de coder donc si l'on garde bien le contrôle sur ce qu'on encode mais on conserve aussi la subjectivité du chercheur dans la construction des nœuds et dans l'attribution des extraits de textes à ces différents nœuds. Ainsi, NVivo, bien qu'étant un outil désincarné, ne permet en aucun cas d'objectiviser l'analyse, son principal intérêt étant de pouvoir l'étendre à un plus grand nombre de cas et de l'approfondir en testant par exemple un plus grand nombre de corrélations.

2.2 L'analyse avec NVivo pas à pas

Dans cette dernière sous-partie, nous proposons de présenter étape par étape comment utiliser le logiciel. Dans la mesure où le codage est une opération relativement longue, il est possible de commencer à coder sans avoir terminé de réaliser l'intégralité des entretiens mais il faut alors s'attendre à ce que les premiers nœuds identifiés lors du codage influencent la conduite des derniers entretiens.

La première étape à laquelle il est conseillé de procéder, bien qu'elle demeure facultative, est le nettoyage des retranscriptions d'entretien. Concrètement, dans la mesure où plusieurs mois peuvent s'écouler entre le premier et le dernier entretien, il s'agit d'homogénéiser les

différentes retranscriptions¹⁴. On peut aussi à cette étape numéroter les questions/réponses au sein de chaque entretien : au moment de la rédaction, cela permettra de référencer plus facilement l'endroit d'où on extrait un verbatim. On peut par exemple utiliser un « E » majuscule pour renvoyer à l'enquêteur/trice et un « e » minuscule pour renvoyer à l'enquêté.e, ou encore un « Q » pour renvoyer aux questions et relances et une « R » pour renvoyer aux réponses, et ensuite numéroter les questions et relances et les réponses. Il peut être préférable de numéroter toutes les interventions de l'enquêteur/trice et non seulement les plus importantes, par exemple en numérotant les « ah ? » et les « hum », dans la mesure où la numérotation vise à se référer plus facilement à la retranscription au moment d'une citation.

On obtient ainsi le schéma suivant :

E1 : comment vous en êtes arrivé là ?¹⁵
e1 : alors c'est une longue histoire euh, un peu euh, peut-être, est-ce que vous voulez que je raconte euh... ?
E2 : allez-y...
e2 : alors en fait, hum, mon précédent poste j'étais à [X] et du coup ça s'est pas bien passé [...]

L'inconvénient de ce codage est que, lorsque l'on souhaite citer un extrait d'entretien, il faut toujours indiquer explicitement au préalable à quel entretien on se réfère. On peut donc faire apparaître le code de l'entretien dans la numérotation. Par exemple si l'on a choisi d'anonymiser en attribuant une lettre à chaque entretien, par exemple pour l'entretien A¹⁶, on obtient le schéma suivant :

[AQ1] : comment vous en êtes arrivée là ?
[AR1] : alors c'est une longue histoire euh, un peu euh, peut-être, est-ce que vous voulez que je raconte euh... ?
[AQ2] : allez-y...
[AR2] : alors en fait, hum, mon précédent poste j'étais à [X] et du coup ça s'est pas bien passé [...]

¹⁴ De manière générale, comme le souligne Abrial (2019), la chercheuse ou le chercheur doit garder à l'esprit l'importance de l'homogénéité des textes analysés dès l'étape de la retranscription : choisit-on de retranscrire les hésitations et comment ? choisit-on de retranscrire les silences et comment ? préfère-t-on indiquer entre crochets dans la retranscription ce qui semble traduire les émotions de l'enquêté.e, par exemple : [s'agite], ou ne pas en tenir compte ? etc.

¹⁵ On souligne que la formulation de la question est intéressante dans la mesure où le « là » final peut renvoyer à toutes les sphères de la vie de l'enquêté.e qui en conséquence choisit la sphère pour laquelle il répond : professionnelle, familiale ou autre.

¹⁶ L'attribution d'une lettre efface complètement l'identité de l'enquêté.e, on peut donc aussi choisir de conserver certaines caractéristiques, par exemple, stratégie souvent utilisée, en transformant le prénom réel de l'enquêté.e en un prénom respectant son genre, sa génération, son origine ethnique ou tout autre élément jugé pertinent pour l'enquête. Par exemple Carmen peut devenir Soledad et Paul devenir Jacques.

Plusieurs signes de ponctuation et éléments de typographie peuvent faire ressortir la numérotation afin qu'elle ne soit pas confondue avec un autre élément de l'entretien : utilisation du gras, de crochets ou autres.

La deuxième étape consiste à créer le fichier sur lequel on va travailler dans le logiciel. Ce dernier aura l'extension « .nvp ». Il peut tout à fait être copié/collé sur un appareil qui ne disposerait pas du logiciel sans être altéré, mais il sera alors impossible de l'ouvrir ou le modifier sur l'appareil en question. Ce procédé est notamment utile pour réaliser des sauvegardes régulières de l'avancement du codage. Le logiciel propose d'ailleurs toutes les quinze minutes de procéder à un enregistrement automatique du fichier. Cette sauvegarde automatique écrase la version précédente du fichier. Or, il peut être parfois utile, à certaines étapes-clés du codage, par exemple une fois que toutes les retranscriptions d'entretiens ont été codées et avant de procéder à une éventuelle réorganisation des nœuds, de conserver la version initiale du fichier. Dans ce cas, comme pour de nombreux logiciels il suffit d'enregistrer le nouveau fichier sous un autre nom, par exemple en y adjoignant la date d'enregistrement ou le numéro de version (V1, V2, V3, etc.¹⁷).

Dans le prolongement de cette volonté de suivre pas à pas la démarche méthodologique de codage dans NVivo, le logiciel propose de créer des « mémos », concrètement des pages de textes qui renvoient à ce que, dans un traitement manuel, on noterait sur un calepin en parallèle de l'analyse. Ces mémos peuvent être rattachés soit à un nœud soit à un entretien. Dans le premier cas on peut vouloir préciser comment le nœud évolue au fur et à mesure de l'avancement du codage qui amène à des changements de contenu du nœud, comme dans l'exemple précédemment développé sur le travail, en justifiant pourquoi ces choix d'ajustement sont opérés¹⁸ ou encore des premiers éléments d'analyse intuitive, par exemple le fait que le nœud « travail » semble corrélé au nœud « famille ». L'intérêt est alors que lorsque cette corrélation apparaît sur les deux ou trois premiers entretiens codés il est encore trop tôt pour réorganiser les nœuds, par exemple en créant un nœud « conciliation travail/famille », dans la mesure où il n'est pas encore établi que la corrélation soit suffisamment importante pour constituer un nouveau code. Dans le second cas, toujours de manière intuitive, on peut noter les impressions laissées au fur et à mesure du codage, ces impressions peuvent porter sur l'enquêté.e, par exemple « enquêté développant spontanément ses réponses », elles peuvent apporter ou compléter les indications émotionnelles si celles-ci ont été notés au moment de la retranscription, par exemple « l'enquêtée semble sûre d'elle, comme le montrent la longueur des réponses développées, l'absence de « euh », sa rapidité à répondre », ou porter sur tout autre élément jugé utile au moment du codage, sur tout élément qui semble important sur le moment, même si l'on n'est pas certain.e que cette importance perdure dans l'analyse finale.

Une fois les retranscriptions nettoyées, il convient de les importer dans le logiciel. De manière générale, NVivo est plutôt intuitif et l'exploration des différents onglets mène rapidement à

¹⁷ En se méfiant des « version finale » qui tendent rapidement à se transformer en « version finale finale » puis en « version finale finale la vraie », etc.

¹⁸ Ces éléments sont particulièrement utiles à des fins d'explication de la méthode au moment de la rédaction.

trouver ce que l'on cherche, du moins pour les opérations de base. Sur la figure 2 on voit ainsi dans le ruban en haut de l'écran les différents onglets disponibles, notamment l'onglet « requête » évoqué ci-dessous que l'on retrouve dans le cadre en bas à gauche. Dans ce même cadre se trouve les différentes options évoquées ci-dessous : les mémos, nœuds, collections et catégories. La deuxième colonne à gauche présente la liste des sources importées sur le logiciel ; au centre se trouve l'entretien sur lequel on travaille et à droite les bandes colorées indiquant le codage¹⁹.

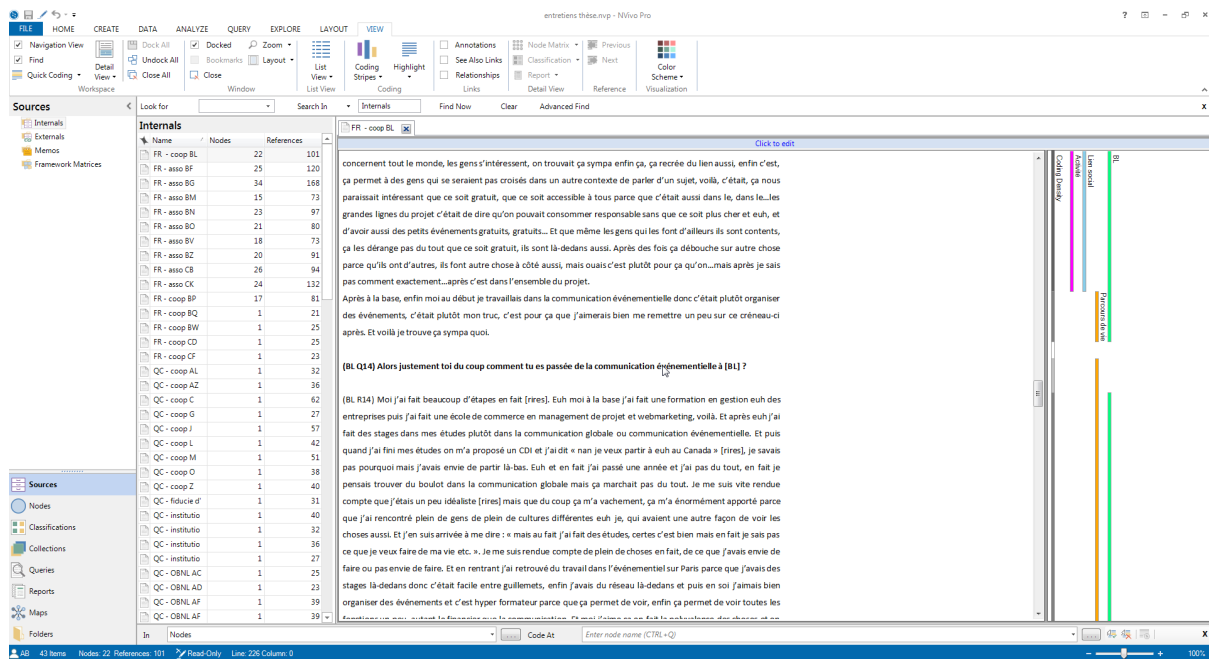


Figure 2 : Présentation générale du logiciel

Les intitulés des entretiens peuvent être plus ou moins explicites, par exemple on peut choisir de faire ressortir le profil socioculturel des enquêté.e.s en choisissant H/F/A²⁰ pour renvoyer au genre, en indiquant l'âge de l'enquêté.e et sa ville d'origine, par exemple : F 50 Grenoble. La première opération à effectuer est d'établir ce que le logiciel nomme les « cas ». Chaque enquêté.e correspond à un cas²¹, le but du cas est, une fois le codage terminé d'identifier quel.le enquêté.e dit quoi dans l'analyse, notamment pour cerner si certains éléments reviennent plus fréquemment chez un certain profil d'enquêté.e.

Une fois les textes importés sur NVivo il est possible de les organiser en les classant par sous-ensembles ou « collections ». Ces ensembles peuvent renvoyer aux lieux dans lesquels les entretiens ont été menés, par exemple Grenoble et Lyon ou Québec et France²², ils peuvent

¹⁹ Il est possible de modifier la taille des différentes colonnes pour plus de confort visuel.

²⁰ Homme/Femme/Autre.

²¹ De la même manière quand on travaille sur des retranscriptions de focus groupe chaque participant.e se voit attribuer un cas

²² Quand on utilise plusieurs types de sources on peut aussi les faire apparaître dans les ensembles, par exemple si l'on mobilise à la fois des retranscriptions d'entretiens et des rapports. Plus largement, le logiciel permet de faire apparaître le même type d'informations à différents endroits, le mieux étant

renvoyer à la nature l'organisation d'appartenance de l'individu si celle-ci est supposée avoir une importance pour l'analyse, par exemple coopération ou association, ils peuvent encore renvoyer à une phase de l'enquête quand le protocole prévoit plusieurs vagues d'entretiens, ou à tout autre élément *a priori* pertinent. Un autre type de classement qu'il est possible d'effectuer est celui par « caractéristiques ». Ces dernières renvoient tant au profil de l'enquêté.e, par exemple son âge, son genre, sa place dans l'organisation (est-il ou elle bénévole ou salarié.e ?) qu'à l'entretien en lui-même, par exemple sa durée, ou s'il a été réalisé à distance par téléphone ou face-à-face.

La deuxième grande étape, que nous avons déjà évoquée, est le codage lui-même, procédé de création de « nœuds » et d'attribution d'extraits de texte à ces « nœuds ». Le fait d'avoir créé des caractéristiques au préalable permet d'étudier séparément les sous-groupes créés, en se concentrant par exemple uniquement sur les bénévoles.

Dans un troisième temps vient l'analyse à proprement parler. NVivo propose plusieurs fonctionnalités que nous ne recenserons pas toutes ici en nous contentant des plus courantes. D'abord, tout comme dans le cadre d'une analyse quantitative, le logiciel permet de réaliser un tri à plat grâce à l'onglet « requête » : il est en effet possible de lui faire compter les mots qui reviennent le plus souvent dans tout ou partie du corpus importé, en excluant les mots non-signifiants tels que les déterminants, les pronoms etc. Pour éviter de faire compter au logiciel les mots de l'enquêteur il faut sélectionner les cas déterminés avant le codage. Cela permet, le cas échéant, d'identifier d'éventuels nœuds qui auraient été ignorés au moment du codage. Pour Camille Duthy « il est indispensable de réaliser un premier tri à plat qui permet une familiarisation avec le corpus. En somme, de répondre à la question : tout enquêté confondu, de quoi traitent ces entretiens ? Pour répondre à cette question, nous nous appuyerons sur trois extractions : une première mettant en lien les nœuds d'abord avec le nombre de références qui y sont encodés (combien de références sont associées à chaque nœud), puis avec le nombre de sources codées (combien d'entretiens mobilisent ce nœud) et enfin avec le pourcentage d'encodage (sur l'ensemble des nœuds, quel pourcentage d'occupation représente un nœud). Ces différents types de données ne disent évidemment pas la même chose : il s'agit de voir d'abord quels sont les thèmes qui reviennent le plus fréquemment dans les discours, puis ceux qui sont abordés par le plus d'enquêtés et enfin ceux dont on parle le plus longtemps » (n.p., p. 48-49). Une fois la requête effectuée, si le résultat est pertinent, il faut penser à l'enregistrer car elle ne l'est pas par défaut.

De plus, NVivo permet, toujours grâce à l'onglet « requête », d'identifier les cooccurrences sous forme de matrice : quels sont les thèmes qui sont fréquemment évoqués ensemble ? Le travail est-il toujours abordé en rapport avec le sens que lui attribue l'enquêté.e ? L'identification des cooccurrences s'inscrit tant dans une démarche hypothético-déductive où l'on vérifiera l'existence de cooccurrences pré-identifiées, que dans une démarche inductive qui fera ressortir des cooccurrences du corpus indépendamment d'hypothèses issues de la

peut-être de choisir l'endroit le plus intuitif ou pratique pour l'analyse. Dans cet exemple précis des lieux d'entretien on peut par exemple faire apparaître le pays où s'est déroulé l'entretien par une abréviation dans son intitulé, par exemple en utilisant des intitulés du type « QC – A », « QC – B », « Fr – C » etc.

théorie, ou encore dans une démarche à mi-chemin de l'une et de l'autre. Mais au-delà d'un croisement des codes entre eux, il est également possible de croiser les codes avec les cas, les caractéristiques et les collections. Ces différents croisements permettent de savoir par exemple si le genre impacte les thématiques abordées dans l'entretien ou si le fait qu'un entretien soit particulièrement long implique que certains thèmes qui n'apparaissent pas dans les entretiens courts sont abordés.

Enfin, NVivo permet de visualiser les résultats sous formes de graphes et schémas ou, selon la préférence de l'utilisateur/trice, d'exporter les résultats vers Excel pour ensuite les modéliser.

Ainsi, NVivo propose une diversité de manières d'intégrer de l'information associée à un entretien : si le codage en tant que tel est la plus intuitive, ou du moins celle qui mène le plus souvent à choisir l'utilisation du logiciel, la mise en place de cas, de collections et de caractéristiques mène à affiner l'analyse. L'utilisateur/trice demeure libre son choix d'intégrer l'information d'une manière plutôt que d'une autre.

CONCLUSION

Les récits de vie sont une méthodologie fortement ancrée dans la sociologie empirique, mais cette méthode offre des perspectives intéressantes pour d'autres disciplines car elle permet de mieux comprendre les interactions entre l'individu, le groupe et la société par la façon dont l'individu se raconte. Aujourd'hui, les occasions de se raconter se multiplient offrant aux chercheurs en sciences humaines et sociales un matériau riche à explorer. Dans ce chapitre nous avons exposé et discuté les fondements, les apports et les usages de cette méthodologie, pratiquée dans le cadre d'un entretien de vive voix.

Dans un second temps, le traitement des données qualitatives s'opère avec des logiciels permettant le traitement d'un grand nombre de mots, de textes, d'images. Nous avons présenté comme traiter des récits de vie avec de tels logiciels, et en particulier le logiciel NVivo. Le logiciel ne substitue pas à l'interprétation et la discussion du chercheur, il est simplement une aide dans l'organisation et la gestion du corpus. Si l'apprentissage du fonctionnement du logiciel constitue un coût d'entrée notable, il permet de traiter qualitativement des corpus plus vastes et plus en profondeur que l'analyse manuelle.

BIBLIOGRAPHIE

Abrial, S. (2019). "Stratégies de codage avec un logiciel CAQDAS". Support de cours, Pacte, Ecole doctorale SHPT, UGA.

Bernard, M.-C. (2014, janvier). La « présentation de soi » : cadre pour aborder l'analyse de récits de vie. *¿ Interrogations ? L'approche biographique*, (17). Disponible sur : <http://revue-interrogations.org/-No17-L-approche-biographique->

Bertaux, D. (2014). *L'enquête et ses méthodes. Le récit de vie*. Paris : Armand Colin, coll. « 128 », (1^{ère} éd. 1997).

- Bourdieu, P. (1986). L'illusion biographique, *Actes de la recherche en sciences sociales*, 62-63, p. 69-72. Disponible sur : https://www.persee.fr/doc/arss_0335-5322_1986_num_62_1_2317
- Brun, P. (2003, novembre). Le récit de vie dans les sciences sociales, *Revue Quart Monde*, 4(188). Disponible sur : <https://www.revue-quartmonde.org/2088>
- Chanfrault Duchet, M. F. (1987). *Cahier de recherche sociologique*, 5(2), 1987, p.11-28.
- Carcassonne, M. (1983). Les notions de médiation et de mimesis chez Paul Ricœur : présentation et commentaires, *Hermès*, (22), p. 53-56. Disponible sur : http://documents.irevues.inist.fr/bitstream/handle/2042/14948/HERMES_1998_22_53.pdf;jsessionid=DC2A919A587CD234246C205F78DBAD3C?sequence=1
- Duthy, C. (non publié). Produire des données. Méthodologie et terrain d'enquête, *Le bien-être à la lumière de la notion d'individuation. Sociologie des solos dans la ville*, thèse de doctorat. Grenoble : Université Grenoble Alpes.
- Giraud, F. ; Raynayd, A. ; Saunier, E. (2014, janvier). Principes, enjeux et usages de la méthode biographique en sociologie, *¿ Interrogations ?*, *L'approche biographique*, (17). Disponible sur : <http://revue-interrogations.org/-No17-L-approche-biographique->
- Hutchison A.J., Johnston L.H. et Breckon J.D. (2010). Using QSR-NVivo to facilitate the development of a grounded theory project: An account of a worked example, *International Journal of Social Research Methodology*, 13(4), p. 283–302.
- Point, S. et Voynet Fourboul, D. (2006). Le codage à visée théorique, *Recherche et applications en marketing*, 21(4).
- Pruvost, G. (2010), Récit de vie, dans Paugam, S., *Les 100 mots de la sociologie*. Paris : PUF. Disponible sur : <http://sociologie.revues.org/671?lang=en>
- Ricœur, P. (1984). *Temps et récit*, vol. II, *La configuration dans le récit de fiction*. Paris : Seuil, coll. « Points Essais ».
- Vanini de Carlo, K. (2014, janvier). Se dire e(s)t devenir - La recherche biographique comme choix épistémologique, *¿ Interrogations ?*, *L'approche biographique*, (17). Disponible sur : <http://revue-interrogations.org/-No17-L-approche-biographique->
- Veith B. (2004, octobre). De la portée des récits de vie dans l'analyse des processus globaux, *Bulletin de méthodologie sociologique*, p. 49-61.

AUTEURS

Amélie Artis

Maître de Conférences en économie
Univ. Grenoble Alpes, Sciences Po Grenoble, CNRS, PACTE, F-38000 Grenoble, France

Anais Bovet

Doctorante en sociologie
Univ. Grenoble Alpes, Sciences Po Grenoble, CNRS, PACTE, F-38000 Grenoble, France

A method for analyzing life histories using CAQDAS software

ABSTRACT

Life stories are a methodology strongly rooted in empirical sociology, but this method offers interesting perspectives for other disciplines because it allows a better understanding of the interactions between the individual, the group and society through the way the individual tells his or her story. In addition, the many advances made by textual data analysis software to handle large amounts of data are encouraging researchers to use life stories. In this chapter, we present the foundations of the life story method. Then in a second part, we explain its treatment with the NVivo text data analysis software. Our approach aims to answer the questions: “Why this method? Why this software?” and “How to do it? How to use it?”

KEYWORDS: life stories, textual data analysis

INTRODUCTION

Qualitative data research has made significant progress through the use of text data analysis software. Like the quantitative ones, these software programs process large amounts of data to help researchers improve their understanding of social phenomena. At the same time, the development of social networks, mainly personal and professional, encourages individuals to tell their stories. How can we grasp the richness of individuals' speeches?

The manuals on research methodology in the humanities and social sciences explain the art of directional, semi-directive or non-directive interviewing by explaining the uses of these interviews, the constitution of interview grids, how to conduct the interview and the limits of this methodology. However, the life story method remains poorly developed outside empirical sociology, although it has significant potential to understand the behaviour of individuals in other disciplines of the humanities and social sciences. Similarly, the practice of life stories and their processing assisted by text data processing software remains poorly addressed in methodological works.

This chapter aims to respond to this pitfall by explaining the foundations, uses and methods of life histories on the one hand, and on the other hand, by presenting their processing by text data analysis software, NVivo in particular. In the first part, we will lay the foundations of the life story method. Then in a second, part we will present its treatment with the NVivo text data analysis software. Our approach aims to answer the questions “Why this method? Why this software?” and “How to do it? How to use them?”

1. LIFE STORIES: A CHOICE TO MANAGE A RESEARCH DESIGN

The life story is a methodological choice as well as an epistemological one. In an inductive process, life stories can reveal repetitive behaviours, triggering factors and interactions that produce similar reactions in order to discover new avenues of research. In a deductive approach, they make it possible to validate hypotheses constructed from the lessons of the scientific literature.

In the first part, we will present the origin of this method, its specificities and limitations. Then, in a second step, we will detail the empirical protocol to make life stories.

1.1 Reveal the meaning of the action

The life story method emerged in contemporary empirical sociology in Chicago in the 1920s and 1930s when students, encouraged by teachers, began working on life stories (Bertaux, 2014). Its origin in sociology can be explained by the fact that "the life story makes it possible (...) to analyse how the symbolic relationship is perceived, felt, lived and theorised in a story, which, through identity play, links the individual to the group, to the community and more broadly to society" (Chanfrault Duchet, 1987).

According to Geneviève Pruvost, the biographical method is based on "a particular type of interview since someone is asked to recall their life and tell their own experience. The system is simple: it is not based on a set of questions/answers based on an interview grid, but on the statement of an initial instruction that invites the narrator to recount the chronological totality of his or her life or part of it, depending on the objective pursued by the interviewer. The sociologist can make reminders, ask questions, but he must ensure that the interview follows the path chosen by the narrator. The sociological interest of the life story lies in this subjective integration: it is a question of grasping the logic of action according to the very meaning that the player gives to his or her trajectory. Far from isolating cases, the life story method makes it possible to locate the network in which the narrator positions himself or herself and to inscribe social phenomena in a chain of causes and effects. The life story helps to highlight processes" (Pruvost, 2010).

Bertaux (2014) identifies three objects that are commonly understood from life stories. First of all, there are the social worlds that can, for example, develop around a professional activity or an unpaid activity. Thus, "the central hypothesis of the ethnosociological perspective is that the logics that govern the whole of a social or mesocosmic world are also at work in each of the microcosms that compose it: by observing in depth only one, or better some of them, and if we succeed in identifying their situation logics, the mechanisms that generate conduct, the processes of reproduction and transformation, we should be able to understand at least some of the social logics of mesocosm itself" (Bertaux, 2014, p. 17). Then, the life stories allow us to be close to what Bertaux calls "the categories of situation" (p. 18), just like single mothers; "This situation is social insofar as it generates, for all those who find themselves there, more or less the same constraints, the same tensions, the same logics (of situation, therefore), a pressure towards logics of action responding to the situation and its tensions" (Bertaux, 2014) p. 18). In these first two objects we find the idea of repetition already mentioned. Finally, the third type of object for which life stories are frequently mobilized is social trajectories. According

to Bertaux, however, there are too many contingencies in a social trajectory for it to be understood through a life story, but these stories could on the other hand refer to a social world.

This method has several specific features. First, it is part of a self-reflective process: during the exchange, the narrator will construct the narrative of his or her life (or part of it), the person interviewed/the narrator has the ability to analyze himself or herself and step back on his or her journey, while formulating a chronological coherence and meaning to the experience. The life story is a way of telling one's story, of presenting oneself to others, and it produces a "narrative identity" (or even a "biographical illusion" for Bourdieu). Through this narrative identity, individuals can be recognized and oriented in their social and professional lives. Then, it participates in the revelation of meaning. Thus Bertaux (2014) distinguishes three orders of reality: i) the historical-empirical reality, that is to say the one actually experienced and acted upon, ii) the biographical journey, or the retrospective thought of the biographical journey, and finally iii) the discursive reality that is produced during the interview.

This method fosters debates about objectivity, or the opposite of the researcher's subjectivity. Indeed, a life story is co-constructed between the narrator and the listener, because it is produced in a context of dialogue in interaction with the researcher that would report to the objectivity of the data and analysis. For example, for Demazière and Dubar (in Bernard, 2014), speech is not the support of the representations already formed but it builds them, the interview which aims to collect life stories is a construction process. Similarly, according to Patrick Brun, Pierre Bourdieu also recognizes the importance of the researcher's relationship with the author of the story.

As an extension of this construction process, from an interactionist perspective, the researcher cannot claim to be neutral and fade during an interview, he is a co-constructor of meaning through the questions he asks, his prompts (Bernard, 2014). By the simple presence of the researcher, and sometimes the orientation of his follow-up questions or his posture, the respondent reacts in a singular way, risking omitting or exaggerating certain points. This relationship can therefore impact the story in its content and form. However, it is first and foremost "the narrator who knows the outcome of his story, [who] selects the elements he considers relevant to understand its meaning" (Bernard, 2014, p. 2). Through this process of constructing the narrative, two types of elements are sorted; thus, "the narratives testify to the ability of social actors to recompose the past (time) and to make mnemonic choices (memory)" (Bernard, 2014, p. 2).

Finally, the respondent could be renamed "the narrator" as soon as his or her words are no longer considered as data but as a constructed narrative, hence the qualification of the interview to collect life stories as a "narrative interview" by Bertaux (2014, p. 10). Thus, "Ricoeur analyses the story as a 'plotting': by telling the story of his or her life, an individual reconfigures his temporal experience, initially confused, around a plot, which transforms the events lived into a story. In addition, he is led to select from among the many events those that seem 'significant' to him" (Giraud et al., 2014, p. 4). This construction thus refers to mimesis II, which is defined as "the moment of the actual plotting by the author of the text. The episodic dimension of the narrative draws narrative time from the linear representation side: the episodes follow each other in accordance with the irreversible order of time common to physical and human events. The configuring dimension, in turn, has temporal traits that are

the opposite of those of the episodic dimension: "the configuring arrangement transforms the succession of events into a significant totality that is the correlate of assembling the events and makes history allow itself to be followed. Thanks to this reflexive act, the story can be translated into a 'thought.' Moreover, 'the configuration of the plot imposes the meaning of an end point' [...] following the incidents indefinitely. By reading the end in the beginning and the beginning in the end, we thus learn to read time itself in reverse, as the recapitulation of the initial conditions of a course of action in its terminal consequences" (Ricoeur, 1983, p. 130-31). Thus, mimesis II transfigures the upstream of the text downstream by its configuration power. It refers to the temporality of the narrative (Carcassonne, 1984, p. 54). More directly in Ricoeur's terms, "plotting has been defined, at the most formal level, as an integrative dynamic that makes a single and complete story from a variety of incidents, in other words, transforms this diversity into a single and complete story" (Ricoeur, 1984, p. 18).

Although admitting its constructed character, the life story as an inquiry process raises questions.

One of the limitations highlighted by the skeptics of the biographical method is characterized by the risk of subjectivity of the researcher in the treatment of the material. Researchers tend to be as objective as possible, but all involve their subjectivity in the way they question, analyze and interpret a subject's story. However, this subjectivity could be exacerbated following a qualitative interview, since it is first and foremost a human encounter. The major risk would be that the researcher, thinking he has the truth, interprets this interview according to his judgment that he thinks is the truest, and risks missing the "truth" of the field. This is all the more the case in a deductive approach where certain hypotheses are already explicitly stated and can thus reduce the interpretation of the identified elements of meaning to those already identified in the scientific literature or, more radically, lead to ignoring certain elements of meaning because they do not fit within the theoretical framework previously stated.

Finally, "the material is eminently qualitative. However, the boundary with quantitative methods is not always so clear," says Blandine Veit (2004). Indeed, it is preferable with such an approach to borrow from several methods to analyze the specific material that is given to us to see with life stories. Therefore, the use of quantitative methods, discourse analysis, and other disciplines such as literature, linguistics and philosophy is not excluded. It also seems relevant to establish timetables, as well as to compare histories from the perspective of individual or collective logics (Veit, 2004). Similarly, the mobilization of several types of sources or media can make it possible to come closer to empirical reality when it is the latter that we are more interested in than the respondent's perception of it, or to explore the different ways in which an individual can tell his story according to the context of enunciation. Thus, for example, the analysis of profiles on social networks can complement the individual's self-report to the interviewer with another form of story intended for another type of audience.

1.2. The art of life storytelling or precise methodology for collecting data

After this brief review of the research positioning implied by the life stories, the objective of this second sub-section is to more concretely present some tips to mobilize this method.

First, data collection most often corresponds to an interview situation. In the case of the respondent, oral forms may be preferred; there is no point in speaking "like a book," it would rather be likely to disturb or even destabilize the respondent, depending on the level of language he or she is used to using. Even if the interview of life stories is a situation of particular interaction through its modalities, it remains an oral exchange between two "humans." For example, one might prefer formulas such as "could you tell me more about...?" over "would you be able to tell me more about...?" If the most neutral option is perhaps to adapt to the respondent's language level so that this point does not constitute an obstacle to the interview, in the absence of knowledge of this characteristic of the respondent, common oral language can be used.

Since the interview as it is most often conducted in a research context is a situation of imbalance between an investigator who is supposed to intervene little and only ask questions and a respondent who is invited to speak at greater or much greater length to answer the questions, it is preferable to avoid anything that could add to the discomfort due to the particularity of the situation. The difficulty is all the more important in the context of the biographical approach where the respondent is not necessarily questioned as an expert on a subject he or she would master or on a context in which he or she evolves, but much more directly on himself or herself.

In the specific case of interviews to collect life histories, the wording "tell me..." is particularly relevant in that it invites the interviewee to tell the story (Bernard, 2014).

In response to the limitations of the biographical interview mentioned above, it is advisable to use formulas that are as neutral as possible to relaunch or prompt the respondent, but without forgetting, once again, that the two interlocutors are human and therefore subject to a certain number of risks. Depending on his or her concentration, mental availability or stage of research, the interviewer will, for example, perform more or less well in his or her interview - it being understood that the term "performance" applied to an interview conducted as part of a research project refers to the interviewer's ability to remain neutral while making the respondent speak. Similarly, some respondents are more or less inclined to digressions, short or evasive answers, or may be more or less available or annoyed by the interview situation... while having previously accepted it.

We propose here some recovery strategies that seem to be applicable regardless of the theme of the interview, while stressing that a good strategy is a strategy that works in the context of a particular research. First, the strategy that probably appears to be the most obvious *a priori* for an investigator who would begin his or her research is to anticipate in advance the follow-up sentences and note them on the interview guide. This makes it possible to anticipate possible formulation biases with a relaxed head and to be able to rely on the preliminary work at the time of the interaction, which can be stressful. Second, vague request for clarification forms such as "i.e." or "can you tell me more about this?" are also very easy to mobilize. Finally, repeating the respondent's words may require more concentration and skill: either you bounce off the last words spoken or, if you take notes at the same time as the narrator unfolds his or her story, you can return to something mentioned earlier. However, it is preferable to be careful to use the terms used by the respondent, in order to avoid already leaving his or her own interpretation and sometimes even directly his or her own analytical framework on the

respondent's words, which could then be taken over by the respondent when the investigator's terms are not part of the spontaneous references. In practice, the respondent's words can be repeated as they stand (see Box 1).

Investigated: and so that's when I thought I had to change my job. Investigator: You thought you needed to change jobs... <i>You can also add one of the previous reminder questions:</i> Investigated: and so that's when I thought I had to change my job. Investigator: You thought you needed to change jobs, can you tell me more (about that)?
--

Sometimes a simple "ah?" or "hum?" is enough to relaunch the respondent, without forgetting also that any communication situation implies a non-verbal language; an expression of questioning on the face can thus call the respondent to continue his or her remarks, to detail more the situation that he or she is evoking, to clarify the theme addressed. Similarly, note-taking in front of the respondent at the time of the interview can influence his or her speech rate by encouraging him or her when the interviewer takes notes or by restricting him or her when he or she does not take notes.

Silent times also have an important role as they allow the respondent to reflect on what he or she is going to say or has just said, and they encourage him or her to talk. Being most often unusual in a spontaneous conversation, the respondent will tend to want to fill them by talking. The interviewer must therefore be skillful in the use of silences and know how to handle them without causing the respondent discomfort, which could have an impact on his or her words. Concretely, three to four seconds of silence are unusual enough without becoming embarrassing. This silence also shows that the interviewer is listening to the respondent and does not seek to interrupt or move to the next section mechanically, as it may sometimes be the case during the first interviews conducted as part of a new or first research. Indeed, it usually takes time to run through the interview guide and familiarize yourself with the reactions it generates in order to eventually adapt it and at least use it properly. In summary, the "golden rule of the narrative interview" (Bertaux, 2014, p. 49) is to "encourage the subject to speak, through simple approvals and reminders, with as little interruption as possible" (*ibid.*)

2. PROCESS AND ANALYZE LIFE STORIES WITH TEXT DATA ANALYSIS SOFTWARE

Like software to process large numbers, there are nowadays software to process and analyze qualitative data. These software tools improve the data processing process (Hutchison et al, 2010) by increasing processing efficiency. This software aid does not replace the researcher's interpretation work, but improves the organization and systematic processing of data. We have chosen to present the NVivo software, which is widely used in the scientific community, in the particular case of life stories. Indeed, this software can have other uses for the researcher such as image processing, but also analysis of a literature review. First, we will recall the foundations and uses of textual analysis, then we will present the use of software in the processing of life histories.

2.1 Text analysis using CAQDAS software

Before starting treatment, the researcher must ask himself several questions. The first concerns the consideration of temporality. There are several types of references to time used by respondents: temporality as such, marked by formulas such as "at the time," memory with "I remember" and retrodiction which consists of making sense of a past element at the time of the interview, with formulations such as "I am talking about it now, I did not see it before" (Bernard, 2014). The second is linked to data; in a structuralist approach, the researcher considers the set of texts as homogeneous, as a set of words to be analyzed, whereas in the interactionist sociolinguistic approach, the transcript of the interview is considered heterogeneous, since the interview does not produce raw data but narrative productions with a beginning, a middle or an end (Bernard, 2014). However, there is nothing to prevent the two approaches from being tested on the same set of texts to compare them. Once these epistemological decisions have been made, the work of cleaning and processing the data can begin.

Textual analysis first consists of identifying, classifying and quantifying the features of a corpus. This processing is first of all descriptive by identifying the topics covered in the corpus, and by producing a frequent vocabulary count to arrive at a statistical analysis. Then, it is a question of making these links appear by crossing the elements between them, and also of studying the thematic concomitances such as: "he talks about his profession, he always immediately addresses the question of money."

The analysis form is the instrument with which the corpus is interrogated. It ranges from the indexing of a few privileged themes in classical thematic analysis to the establishment of a real questionnaire with its groups of questions, scales and numerical indices. Researchers must use a grid of at least three degrees. For example, the theme of "family relationships in your childhood" includes i) family members (father, mother, subject, brother), ii) types of relationships (authority, information, aggressive, supportive, emotional, physical), iii) interrelationships (father-mother, father-subject, father-brother etc.).

To carry out the processing, the NVivo software makes it possible to code texts, interview transcriptions or other texts, i.e. to assign keywords, which are called "nodes" by the software, to extracts of which the length is chosen. But this process itself raises several questions. First, which scale to choose, i.e. which extract size is most relevant for the analysis? Glaser and Strauss (in Point and Voynnet-Fourboul, 2006) suggest not so much to choose *a priori* a scale but rather to divide the text according to the occurrence of a "critical incident" (p. 64) understood as "an idea, or a set of ideas that can be isolated from the rest of the qualitative data and that represents a certain cohesion" (ibid.). More than a mechanical tool that would apply without reflection, the breakdown of critical incidents is partly based on intuition; a precaution to be taken is to keep elements of context in each extract in order to avoid distorting the respondent's proposals at the time of the analysis itself. Each extract can be coded several times, i.e. assigned several nodes. The interest is then to be able to highlight possible co-occurrences between the different codes. The different codings can be stacked without overlapping in terms of the boundaries of the extract. An extract A can end with a sentence that also belongs to the extract B and the beginning of the extract B does not have to go back

to the beginning of the extract A. In Figure 1 for example, on the right of the screenshot, we can see the bands of different colours representing the coded passages.

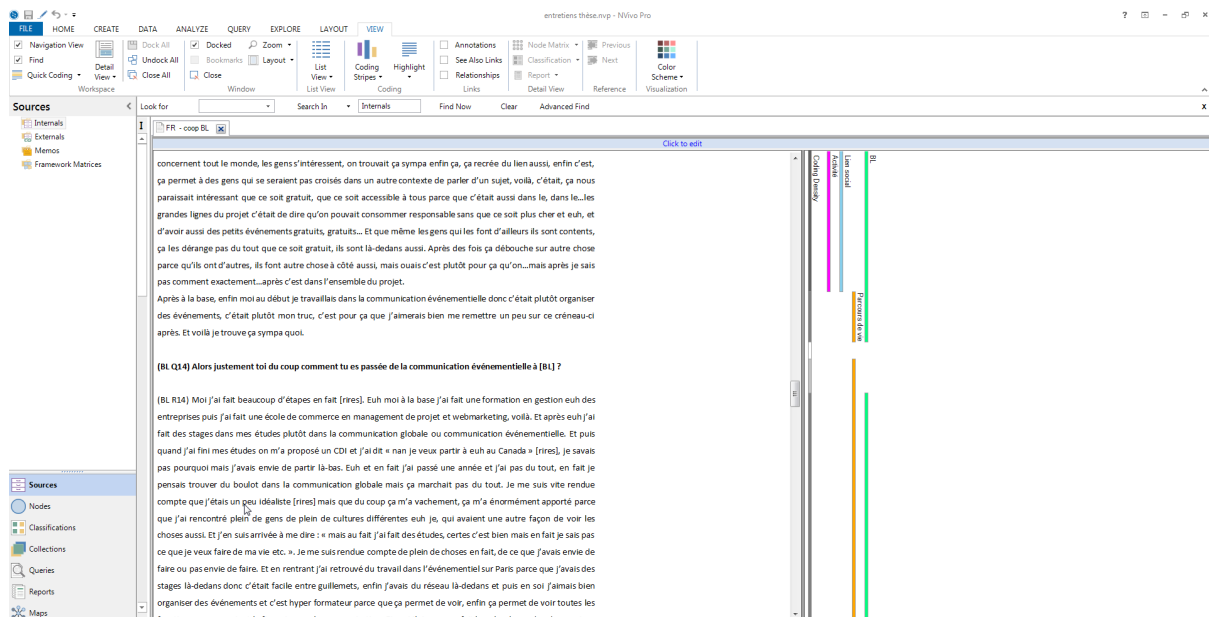


Figure 1 : Overlapping coding on NVivo

Then, what approach should be used to implement these codes? Here we find the two main research approaches at stake: hypothetical-deductive and inductive. In agreement with Point and Voynnet, we recommend adopting a posture halfway between the two, what they call "coding for theoretical purposes" (2006, p. 64) and which consists of "coupling these two approaches, in order to proceed iteratively. It is then a question of encouraging 'back and forth' between the data collected and the researcher's successive interpretations, which we assimilate to an abduction process" (*ibid.*). It is also possible, as the coding progresses, to make adjustments between nodes by creating subcodes for example, but also by specifying or amending the description of a code, which the software allows since a description paragraph is also possible beyond the code title. For example, it may be realized that a node is too large and refers to a diversity of situations that are sometimes contradictory to each other and ultimately make the node irrelevant in terms of analysis, hence the need to subdivide it. Thus the "work" node described as "all the elements on the relationship to work" can be broken down into "job satisfaction" and "dissatisfaction," which in turn can give rise to the nodes "dissatisfaction with material working conditions" and "lack of self-realization." Stéphanie Abrial (2019) advises to rely on several coding strategies that she refers to as thematic, transversal, contextual/emotional and sequential codes. The first refers to elements related to the research; the second to elements *a priori* unrelated to the research but which are likely to be found in any interview, for example questions about money; the third can describe both the conditions under which the interview is conducted, for example the place where it takes place, and the respondent's reactions, for example the laughter or silence; the fourth, finally, reveals the structure of the interview. When the interview guide is rigorously followed, this last strategy leads to the identification of codes that directly reflect it.

Finally, a limitation, or at least a reminder, that deserves to be stressed is that the software only codes what it is asked to code and in the way in which it is asked to code. So if we keep

control over what we encode, we also keep the subjectivity of the researcher in the construction of the nodes and in the allocation of text extracts to these different nodes. Thus, NVivo, although a disembodied tool, in no way makes it possible to objectify the analysis, its main interest being to be able to extend it to a greater number of cases and to deepen it by testing a greater number of correlations, for example.

2.2 Using Nvivo Software step-by-step

In this last sub-section, we propose to present how to use the software step-by-step. Since coding is a relatively long process, it is possible to start coding without completing all the interviews, but it is expected that the first nodes identified during coding will influence the conduct of the last interviews.

The first step that is recommended, although optional, is the cleaning of the interview transcripts. In concrete terms, since several months may elapse between the first and last interview, the aim is to homogenize the various transcriptions. It is also possible at this stage to number the questions/answers within each interview; at the time of writing, this will make it easier to reference the place from which a verbatim is extracted. For example, an "I" can be used to refer to the interviewer and a "R" to refer to the respondent, or a "Q" to refer to questions and reminders and an "A" to refer to answers, and then number the questions and reminders and answers. It may be preferable to number all the investigator's interventions and not just the most important ones, for example by numbering the "ah?" and "hum," since the numbering is intended to make it easier to refer to the transcript at the time of a citation.

This results in the following diagram:

I1: how did you get here? R1: so it's a long story uh, a little bit uh, maybe, do you want me to tell uh...? I2: go ahead... R2: so actually, um, my previous position I was at [X] and so it didn't go well [...]

The disadvantage of this coding is that, when quoting an excerpt from an interview, it is always necessary to indicate explicitly in advance which interview is being referred to. The interview code can therefore be displayed in the numbering. For example, if we have chosen to anonymize by assigning a letter to each interview, for example for interview A, we obtain the following diagram:

[AQ1] : How did you get there? [AA1] : So it's a long story, uh, a little bit, uh, maybe, do you want me to tell uh...? [AQ2] : Go ahead... [AA2] : so actually, um, my previous position I was at [X] and so it didn't go well [...]
--

Several punctuation marks and typographical elements can highlight the numbering so that it is not confused with another element of the interview: use of bold, square brackets or other.

The second step consists in creating the file on which we will work in the software. The latter will have the extension ".nvp". It can be copied/pasted on a device that does not have the software without being altered, but it will then be impossible to open or modify it on the device

in question. This process is particularly useful for making regular backups of the coding progress. The software also offers to automatically save the file every fifteen minutes. This automatic backup overwrites the previous version of the file. However, it may sometimes be useful, at certain key stages of coding, for example once all interview transcripts have been coded and before any reorganisation of nodes is carried out, to keep the initial version of the file. In this case, as with many software programs, it is sufficient to save the new file under another name, for example by adding the registration date or version number (V1, V2, V3, etc.).

As an extension of this desire to follow the methodological approach of coding in NVivo step-by-step, the software proposes the creation of "memos," specifically pages of text that refer to what, in manual processing, would be noted on a notebook in parallel with the analysis. These memos can be attached to either a node or an interview. In the first case, we may want to specify how the node evolves as the coding progresses, which leads to changes in the content of the node, as in the example previously developed on work, by justifying why these adjustment choices are made or the first elements of intuitive analysis, for example the fact that the "work" node seems correlated to the "family" node. The interest is that when this correlation appears on the first two or three coded interviews, it is still too early to reorganize the nodes, for example by creating a "work/family balance" node, since it has not yet been established that the correlation is sufficiently important to constitute a new code. In the second case, always intuitively, we can note the impressions left as the coding progresses. These impressions can relate to the respondent, for example "respondent spontaneously developing his answers," and they can provide or complete emotional indications if they were noted at the time of transcription, for example "the respondent seems confident, as shown by the length of the developed answers, the absence of 'uh,' his speed of response," or cover any other element considered useful or important at the time of coding, even if it is not certain. This importance continues in the final analysis.

Once the transcripts have been cleaned, they must be imported into the software. In general, NVivo is quite intuitive and exploring the different tabs quickly leads to finding what you are looking for, at least for basic operations. In Figure 2, we see in the ribbon at the top of the screen the different tabs available, in particular the "request" tab mentioned below, which is found in the lower left frame. In the same frame are the different options mentioned below: memos, nodes, collections and categories. The second column on the left shows the list of sources imported into the software; in the center is the interview on which we are working and on the right the coloured bands indicating the coding.

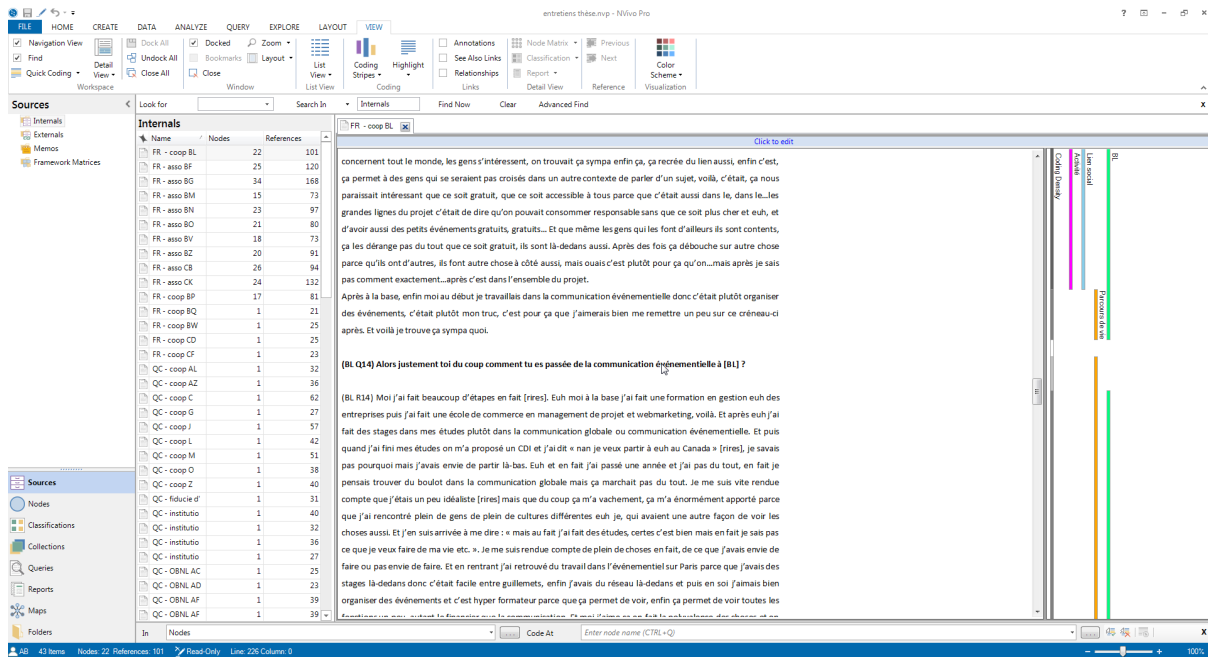


Figure 2: General presentation of the software

The titles of the interviews may be more or less explicit, for example, one can choose to highlight the socio-cultural profile of the respondents by choosing M/W/O to refer to gender, and indicating the respondent's age and city of origin, for example resulting in "W 50 Grenoble."

The first step is to establish what the software calls the "cases." Each respondent corresponds to a case, the purpose of the case is, once the coding is completed, to identify which respondent says what in the analysis, in particular to determine if certain elements return more frequently to a certain respondent profile.

Once the texts have been imported into NVivo, it is possible to organize them by sub-sets or "collections." These sets may refer to the places where the interviews were conducted, for example Grenoble and Lyon or Quebec and France. They may refer to the nature of the individual's organization of belonging if it is supposed to be important for the analysis, for example cooperatives or non-profit organizations. They may also refer to a phase of the survey when the protocol provides for several waves of interviews, or to any other element *a priori* relevant. Another type of classification that can be made is that of "characteristics." The latter refer both to the respondent's profile, for example age, gender, position in the organisation (is he or she a volunteer or paid employee?) and to the interview itself, for example duration or whether it was conducted by telephone or face-to-face.

The second major step, which we have already mentioned, is the coding itself or the process of creating "nodes" and assigning text extracts to these "nodes." Having created characteristics in advance allows the subgroups created to be studied separately, focusing for example only on volunteers.

The third step is the actual analysis. NVivo offers several features, of which we will not list all here, but only the most common ones. First, as in the case of a quantitative analysis, the software allows a flat sorting thanks to the "query" tab; it is indeed possible to make it count

the words that recur most often in all or part of the imported corpus, excluding non-significant words such as determinants, pronouns etc. To avoid having the software count the interviewer's words, select the cases determined before coding. This allows, if necessary, to identify any nodes that may have been ignored at the time of coding. For Camille Duthy "it is essential to carry out a first flat sorting which allows a familiarization with the corpus. In short, to answer the question: what are these interviews about? To answer this question, we will rely on three extractions: a first one linking the nodes first with the number of references encoded there (how many references are associated with each node), then with the number of coded sources (how many interviews mobilize this node) and finally with the encoding percentage (on all nodes, what percentage of occupation represents a node). These different types of data obviously do not say the same thing: it is a question of first seeing which themes are most frequently mentioned in the speeches, then those that are addressed by the most respondents and finally those about which we talk the longest" (n.p., p. 48-49). Once the query has been made, if the result is relevant, you should think about saving it because it is not done by default.

In addition, NVivo allows, still thanks to the "request" tab, to identify co-occurrences in the form of a matrix: which themes are frequently mentioned together? Is the work always approached in relation to the meaning given to it by the respondent? The identification of co-occurrences is part of both a hypothetical deductive approach where the existence of pre-identified co-occurrences will be verified, and an inductive approach that will bring out co-occurrences of the corpus independently of hypotheses derived from theory, or a midway approach to both. But beyond a cross-referencing of codes between them, it is also possible to cross-reference codes with cases, characteristics and collections. These different crossovers make it possible to know, for example, whether gender has an impact on the themes covered in the interview or whether the fact that an interview is particularly long implies that certain themes that do not appear in the short interviews are covered.

Finally, NVivo allows you to view the results in graphs and diagrams or, depending on the user's preference, to export the results to Excel and then model them.

Thus, NVivo offers a variety of ways to integrate information associated with an interview. If the coding as such is the most intuitive, or at least the one that most often leads to choosing the use of the software, the implementation of cases, collections and characteristics leads to a more refined analysis. The user remains free to choose to integrate the information in one way or another.

CONCLUSION

Life stories are a methodology strongly rooted in empirical sociology, but this method offers interesting perspectives for other disciplines because it allows a better understanding of the interactions between the individual, the group and society through the way the individual tells his or her story. Today, opportunities to tell their stories are multiplying, providing researchers in the humanities and social sciences with rich material to explore. In this chapter we have presented and discussed the foundations, contributions and uses of this methodology, which is practiced in an oral interview.

In a second step, qualitative data processing is carried out using software that allows the processing of a large number of words, texts and images. We presented how to deal with life histories with such software, and in particular NVivo software. The software does not replace the researcher's interpretation and discussion; it is simply an aid in the organization and management of the corpus. While learning how to use the software is a notorious cost of entry, it makes it possible to qualitatively treat larger and more in-depth corpuses than manual analysis.

REFERENCES

- Abrial, S. (2019). "Stratégies de codage avec un logiciel CAQDAS". Support de cours, Pacte, Ecole doctorale SHPT, UGA.
- Bernard, M.-C. (2014, January). La « présentation de soi » : cadre pour aborder l'analyse de récits de vie. *¿ Interrogations ? L'approche biographique*, (17). Disponible sur : <http://revue-interrogations.org/-No17-L-approche-biographique->
- Bertaux, D. (2014). *L'enquête et ses méthodes. Le récit de vie*. Paris : Armand Colin, coll. « 128 », (1^{ère} éd. 1997).
- Bourdieu, P. (1986). L'illusion biographique, *Actes de la recherche en sciences sociales*, 62-63, p. 69-72. Disponible sur : https://www.persee.fr/doc/arss_0335-5322_1986_num_62_1_2317
- Brun, P. (2003, November). Le récit de vie dans les sciences sociales, *Revue Quart Monde*, 4(188). Available on : <https://www.revue-quartmonde.org/2088>
- Chanfrault Duchet, M. F. (1987). *Cahier de recherche sociologique*, 5(2), 1987, p.11-28.
- Carcassonne, M. (1983). Les notions de médiation et de mimesis chez Paul Ricœur : présentation et commentaires, *Hermès*, (22), p. 53-56. Available on : http://documents.irevues.inist.fr/bitstream/handle/2042/14948/HERMES_1998_22_53.pdf;jsessionid=DC2A919A587CD234246C205F78DBAD3C?sequence=1
- Duthy, C. (unpublishing). Produire des données. Méthodologie et terrain d'enquête, *Le bien-être à la lumière de la notion d'individuation. Sociologie des solos dans la ville*, thèse de doctorat. Grenoble : Université Grenoble Alpes.
- Giraud, F. ; Raynayd, A. ; Saunier, E. (2014, January). Principes, enjeux et usages de la méthode biographique en sociologie, *¿ Interrogations ? , L'approche biographique*, (17). Disponible sur : <http://revue-interrogations.org/-No17-L-approche-biographique->
- Hutchison A.J., Johnston L.H. et Breckon J.D. (2010). Using QSR-NVivo to facilitate the development of a grounded theory project: An account of a worked example, *International Journal of Social Research Methodology*, 13(4), p. 283–302.

- Point, S. et Voynnet Fourboul, D. (2006). Le codage à visée théorique, *Recherche et applications en marketing*, 21(4).
- Pruvost, G. (2010), Récit de vie, dans Paugam, S., *Les 100 mots de la sociologie*. Paris : PUF. Disponible sur : <http://sociologie.revues.org/671?lang=en>
- Ricœur, P. (1984). *Temps et récit*, vol. II, *La configuration dans le récit de fiction*. Paris : Seuil, coll. « Points Essais ».
- Vanini de Carlo, K. (2014, January). Se dire e(s)t devenir - La recherche biographique comme choix épistémologique, *¿ Interrogations ?*, *L'approche biographique*, (17). Disponible sur : <http://revue-interrogations.org/-No17-L-approche-biographique->
- Veith B. (2004, October). De la portée des récits de vie dans l'analyse des processus globaux, *Bulletin de méthodologie sociologique*, p. 49-61.

AUTHORS

Amélie Artis

Associate Professor in Economics

Univ. Grenoble Alpes, Sciences Po Grenoble, CNRS, PACTE, F-38000 Grenoble, France

Anais Bovet

PhD Candidate in Sociology

Univ. Grenoble Alpes, Sciences Po Grenoble, CNRS, PACTE, F-38000 Grenoble, France

Mobiliser l'empathie dans nos méthodes de recherche : des propositions inspirées du design thinking

RÉSUMÉ

En sciences de Gestion, les émotions ont connu un intérêt grandissant à partir du milieu des années 90. Le rôle des événements organisationnels comme « amorce émotionnelle » a ainsi été théorisé dans l'*affective events theory*. Le chercheur qui souhaite comprendre des comportements dans et de l'entreprise peut prendre en compte cette dimension émotionnelle en étant capable d'empathie pour s'appropriier et comprendre en profondeur ce qui se joue dans l'organisation. Les outils de la méthodologie du Design Thinking (méthode d'innovation centrée sur l'humain, mobilisant pensée analytique et pensée intuitive) sont particulièrement appropriés pour mobiliser l'empathie dans la recherche en Gestion.

MOTS CLES : Empathie, Emotions, Design thinking, Méthodologie, Sense-making

INTRODUCTION

La recherche se nourrit d'expériences multiples, et de leur confrontation peut émerger une meilleure compréhension des phénomènes organisationnels que l'on souhaite étudier. De plus, par son ampleur et sa variété, il est souhaitable que la recherche en gestion ait recours à des méthodologies et des paradigmes variés. Le chercheur doit se faire confiance et accepter parfois de modifier son plan d'analyse primitif, car les conditions d'observation au sein d'une organisation peuvent évoluer, remettant en question les orientations méthodologiques initiales. Être « à l'écoute » du terrain et de ses acteurs, et accepter de faire des aller-retours dialectiques entre objectifs, méthodes et analyses, garantit souvent l'émergence d'une bonne recherche. La capacité du chercheur à saisir les émotions des acteurs rencontrés, à faire preuve d'empathie pour aborder le terrain et les subtilités des situations - rencontrées et racontées - est rarement évoquée dans les ouvrages ou articles sur les méthodologies de recherche. Pourtant, c'est une compétence du chercheur qui devrait être valorisée car, parmi d'autres, elle permet une analyse fine et globale du terrain de recherche. Bien entendu, les psychologues et les anthropologues (entre autres), ont développé beaucoup de méthodologies et d'outils pour capter et mesurer les émotions, dont les chercheurs en Gestion doivent s'inspirer. Le design thinking, méthodologie récente d'innovation basée sur une approche pluridisciplinaire remettant l'utilisateur, son ressenti et ses expériences au cœur de l'analyse, est devenu une source intéressante de captation et compréhension des émotions.

A partir d'une analyse systématique de la littérature sur le design thinking, Micheli et al. (2019) ont proposés neuf attributs pour définir le design thinking : " raisonnement abductif ", " équilibre ", " collaboration ", " outils de conception ", " innovation ", " résolution de problèmes ", " vision systémique ", " tolérer l'échec " et " centré sur l'utilisateur ". Ces neuf attributs, et surtout les outils qui les accompagnent, vont nous amener à proposer une autre manière d'aborder la méthodologie de recherche en sciences de gestion, intégrant l'empathie.

1. L'EMPATHIE : UN CONCEPT À EXPLORER EN SCIENCES DE GESTION

En sciences de Gestion, les émotions ont connu un intérêt grandissant à partir du milieu des années 90. Le rôle des événements organisationnels comme « amorce émotionnelle » a été théorisé dans l'*affective events theory* (Weiss et Cropanzano, 1996).

Depuis, plusieurs théories de l'émotion issues de la psychologie ont été introduites en sciences de Gestion afin « de montrer que 1/ le comportement organisationnel est déterminé en grande partie par l'expérience émotionnelle de l'individu ; 2/ l'environnement organisationnel est une source discrète d'évènements affectifs qui déclenchent ces émotions ; et 3/ les réponses des individus à ces événements déterminent la nature des attitudes et des intentions comportementales subséquentes » (Lajante et Lux, 2018, p.7-8).

Le chercheur qui souhaite comprendre des phénomènes et des comportements dans et de l'entreprise doit prendre en compte cette dimension émotionnelle et donc être capable d'empathie pour s'approprier et comprendre en profondeur ce qui se joue dans l'organisation.

1.1. Définitions

Étymologiquement (dictionnaire le Petit Robert), le mot empathie dérive du radical latin *in-* et *Im-* (*en*) et du groupe suffixal grec *-pathein* (*sentir*) : il exprime la capacité à s'identifier à autrui et à épouser la perception subjective d'autrui.

L'empathie est un concept nomade, encore fragile, et défini par plusieurs courants de pensée. Elle est distincte de la sympathie qui correspond, elle, à une contagion émotionnelle. Ainsi, si l'empathie consiste à comprendre autrui, être en sympathie avec l'autre consiste à se soucier de son bien-être (Wispé 1987). L'empathie est alors « la capacité à se mettre à la place d'une autre personne pour comprendre ses sentiments, ou à se représenter la représentation mentale d'une autre personne » (Boulanger et Lançon, 2006). L'empathie s'exprime alors à travers différents phénomènes comme la projection, l'identification ou l'altruisme, et correspond à une simulation mentale consciente de la subjectivité d'autrui.

L'empathie est un phénomène complexe, au cœur du Soi, en lien avec les émotions et l'expérience (empathie narrative). Pour la comprendre, il faut se référer aux théories psychologiques, phénoménologiques et aux neuro-sciences... mais c'est la philosophie morale écossaise qui fut la première à introduire la notion d'empathie. Boulanger et Lançon (2006, p.498) citent Hume qui observait dans le *Traité de la nature humaine* que « les esprits de tous les hommes sont semblables quant à leurs sentiments et à leurs opérations, et il n'y a

pas d'inclination ressentie par un homme qui ne puisse également affecter tous les autres à un certain degré ». Plus tard, Husserl (1929, 2003) fait de l'empathie le phénomène clé sur la base duquel une intersubjectivité s'établit pour constituer un mode comportemental commun. L'empathie husserlienne est donc avant tout une forme de médiation cognitive.

L'empathie peut alors être comprise non pas comme un mécanisme en soi, mais comme un processus d'analyse et de communication. Selon C. Rogers (1985), l'empathie consiste à saisir avec autant d'exactitude possible les références internes et les composantes émotionnelles d'une autre personne ... et à les comprendre comme si on était cette autre personne.

1.2. Un allié pour comprendre le sens des actions

Et si on intégrait l'empathie à nos méthodes de recherche ? Oui, parce qu'elle est importante pour comprendre le sens des actions et réflexions des acteurs (« *sense-making* »). Et plus particulièrement pour :

- découvrir les émotions qui guident les comportements,
- identifier les bons interlocuteurs,
- guider les efforts d'innovation.

Le modèle théorique du *sense-making* est présenté chez Dervin (1992, 1998) comme une métaphore caractérisée par trois concepts fondamentaux : situation initiale, manque ou discontinuité, aide ou résultats. Ces concepts se traduisent par trois questions qui indiquent le caractère dynamique et itératif du processus de *sense-making* : qu'est-ce qui manque dans votre situation et qui fait en sorte que vous vous arrêtez ; quelles questions avez-vous ou quelles ambiguïtés ressentez-vous ; quel genre d'aide espérez-vous obtenir pour pouvoir continuer ? L'individu est appelé à expérimenter différents types de discontinuités dans une situation prenant place dans un contexte précis (Maurel 2010).

Weick, pour sa part, pose cette question : « comment pouvons-nous savoir ce que nous pensons (ou voulons) avant d'avoir vu (ou compris) ce que nous disons (ou faisons) ? ». Pour lui, cette phrase contient à la fois la recette de la construction du sens (*sense-making*) et de l'acte de s'organiser (*organizing*). Si le sens rend possible l'acte de s'organiser, celui-ci à son tour rend possible le sens : on s'organise à travers et pour le sens, et la création de sens permet aux acteurs organisationnels de partager des modèles mentaux, des valeurs et des croyances, de coordonner des activités et d'agir en interrelation. La réalité est construite par les acteurs (qui « énaquent » leur environnement) à travers les interprétations qu'ils lui donnent. C'est parce que la création de sens est rendue nécessaire par l'ambiguïté inhérente à l'environnement, qu'elle est, selon Weick (2001), une problématique organisationnelle centrale.

D'un point de vue théorique, Weick (1993) définit le processus de création de sens comme une démarche continue à travers laquelle les individus visent à rendre rationnelles pour eux-mêmes et pour les autres les actions et situations qu'ils traversent. Le *sense-making* fait référence à un processus cognitif permettant de saisir le sens d'une situation en lui attribuant une signification.

Weick (1993) propose d'approcher ce processus à l'aide du modèle ESR : « enactment, sélection, rétention ». Dans la phase « d'enactement », l'acteur organisationnel essaie de donner du sens au contexte dans lequel il se trouve. Il cherche à réduire l'équivocité d'une situation, à savoir la présence d'interprétations multiples pour une même situation. La phase de sélection revient à choisir le schéma explicatif acceptable par tous, parmi les différents schémas de pensée construits. Enfin, la phase de rétention vise à garder en mémoire collective les situations rencontrées, ainsi que les interprétations effectuées, pour pouvoir y recourir au moment où une situation similaire se présentera. Le caractère rétrospectif du processus de *sense-making* met en lumière un modèle d'apprentissage organisationnel. En effet, il n'est pas envisageable de donner du sens à un fait ou à une action avant qu'elle ne soit survenue. « Ce n'est seulement qu'après avoir vécu une situation qu'il est possible de regarder en arrière pour construire sa signification (Argyris & Schön, 2001). Les expériences actuelles sont alors comparées aux expériences passées pour créer du sens » (Guilmot et Vas, 2012, p.84).

Le processus de construction du sens est avant tout propre à chaque individu, puisqu'il est marqué par des facteurs cognitifs et affectifs inhérents à l'être humain, qu'il s'agisse de sa manière de percevoir sa réalité ou encore d'analyser et d'interpréter des éléments d'information. Mais ce processus est également contextuel, puisqu'il prend place dans un environnement social, culturel, économique et politique qui caractérise la situation vécue par un individu, un groupe ou une organisation.

Le chercheur qui souhaite appréhender au mieux le sens des actions et réflexions des acteurs d'une organisation, ou trouver les émotions qui guident les comportements, doit faire preuve d'empathie. Pour cela, il peut s'appuyer sur des propositions méthodologiques, dont la plupart sont issues du *design thinking*.

2. PROPOSITIONS MÉTHODOLOGIQUES INSPIRÉES DU DESIGN THINKING

Pour apporter de l'empathie aux méthodologies de recherche en Gestion, il est primordial, avant tout, d'être dans un état d'esprit ouvert et attentif, et être prêt à bousculer un peu les méthodologies classiques pour les enrichir ou les exploiter différemment. Les quatre propositions suivantes permettent d'aborder simplement la méthodologie de recherche avec une résolution d'empathie claire inspirée du *design thinking*. Celui-ci se définit comme une méthode d'innovation centrée sur l'humain, qui se veut une synthèse entre pensée analytique et pensée intuitive. Il permet une double logique de **divergence/convergence** (phase d'intuition et analogique) et une volonté de compréhension pleine et entière des expériences (importance de l'étude terrain, approche **ethnographique**) (Brown, 2008). Le *design thinking* s'utilise en général dans une approche socio-constructiviste.

2.1. Adopter l'état d'esprit du débutant

Parce que nous portons tous avec nous nos expériences, notre vision du monde, notre expertise, nous avons des préjugés lorsque nous démarrons une recherche. Nous souhaitons tous réussir à aborder une problématique sans ces opinions préalables, et laisser le terrain

nous dire ce qu'il en est, mais ce n'est pas simple. Les hypothèses très précises, les stéréotypes peuvent être de fausses pistes qui se révéleront des freins à la compréhension réelle du sens de l'action des personnes observées, parce qu'ils limitent l'empathie. Adopter l'état d'esprit d'un novice, mettre de côté ses préjugés, permet d'aborder la recherche avec des yeux neufs, et d'éviter de passer à côté du sens donné par les acteurs à leurs choix et à leurs actions. Mais comment faire pour adopter cet état d'esprit du novice ?

1. Essayer de ne pas juger. Observer simplement, interviewer sans orientations pouvant amener des jugements de valeurs. La façon dont les questions sont formulées est importante pour rester neutre.
2. Questionner tout... même et surtout ce que l'on croit déjà connaître. Poser des questions pour en apprendre davantage sur la façon dont la personne perçoit le monde. S'imaginer comme un enfant qui veut tout comprendre et demande et répète « pourquoi ».
3. Etre vraiment curieux et s'efforcer de prendre une posture de curiosité surtout dans les situations qui semblent familières ou qui mettent mal à l'aise.
4. Ecouter vraiment. Ne pas penser au prochain thème qui va être abordé, mais se concentrer sur ce que disent les interlocuteurs, et comment ils le disent.

2.2. Observer plus finement avec le QCP (continuum concret – émotionnel)

Pour décoder le plus largement possible un continuum concret (quoi)-émotionnel (pourquoi) dans les réponses apportées ou dans les observations d'actions, il est intéressant de mobiliser le « QCP : quoi, comment, pourquoi ». Ces simples questions permettent de passer de l'observation concrète d'une situation aux émotions les plus abstraites et aux motivations en jeu dans la situation. C'est très utile notamment dans l'analyse d'observation in-situ ou de photos ou d'images. Il faut commencer par des observations concrètes (quoi). Que fait la/les personne(s) observée(s) ou photographiée(s) ? Il convient d'aller dans le détail, d'essayer d'être objectif et de ne pas faire d'hypothèses à ce stade. Puis il faut passer à la compréhension (comment). Comment la personne observée fait ce qu'elle fait ? Cela lui demande-t-il un effort ? Semble-t-elle pressée, contente, en difficulté ? Il est préférable d'utiliser des phrases descriptives, avec des adjectifs précis. Enfin reste la phase d'interprétation (pourquoi). Pourquoi la personne fait ce qu'elle fait, et pourquoi de cette manière ? Cette étape nécessite des suppositions quant à la motivation et les émotions de cette personne pour donner du sens à la situation observée.

2.3. Préparer et mener un entretien différemment

Pour comprendre les pensées, les émotions et les motivations d'une personne, il est utile de l'encourager à une parole riche et complète. Les guides d'entretien semi-directifs que nous utilisons beaucoup en méthodologie qualitative en Gestion sont souvent trop « indicatifs » et biaisent les réponses, ou les ferment trop tôt. Pour interviewer de manière empathique, il faut demander pourquoi, et encore pourquoi ! Même si on pense connaître la réponse, demander aux gens pourquoi ils font ou disent certaines choses peut amener des réponses surprenantes. La conversation naissant d'une question devra continuer aussi longtemps qu'elle semble utile. De plus, il faut éviter d'utiliser le vocable « habituellement » dans les questions posées. Au lieu de cela, il est préférable de demander aux personnes de narrer une situation spécifique

en formulant la question sous la forme « racontez moi la dernière fois que vous ... ». Ainsi, il est intéressant d'encourager les histoires. Qu'elles soient vraies ou non, elles révèlent la façon dont les personnes pensent le monde. Quant aux incohérences entre ce que les gens disent, et ce qu'ils font, elles cachent parfois des enseignements intéressants pour le chercheur. Tout comme les silences, il ne faut pas chercher à les combler car elles illustrent souvent une réflexion plus profonde de la personne sur le thème en cours. Il faut aussi éviter les questions fermées qui n'initient pas d'histoire, et éviter les questions qui pourraient suggérer qu'il y a une bonne réponse.

2.4. Synthétiser en saturant puis en groupant

Pour compléter l'approche empathique d'une recherche, il est souvent pertinent de rendre tangible les pensées, expériences, interviews sous forme de mots-clés ou d'images, bref d'éléments tangibles et visuels propices à la synthèse et à l'émergence de thèmes. Regroupés sur des post-it, l'idée est d'en afficher un maximum (de saturer le mur !) puis de les rassembler par thèmes qui pourront guider les réflexions et l'analyse. Il est ensuite intéressant de creuser chaque thématique en cherchant des connexions avec les autres thèmes ou en approfondissant chaque idée exprimée dans le thème. On retrouve ici la même logique que lors de séance d'idéation en *design thinking*, où le brainstorming permet de diverger puis l'analyse thématique de converger. Pour plus d'efficacité, vous pouvez réunir quelques collègues qui font en même temps que vous le travail de convergence, et/ou qui notent leurs constats après vos regroupements est toujours utile... une façon de tester et discuter un prototype, en quelque sorte !

Au-delà de ces quatre postures méthodologiques, des outils précis peuvent accompagner le chercheur dans son approche empathique du terrain.

3. PROPOSITIONS D'OUTILS POUR INTÉGRER L'EMPATHIE AUX MÉTHODES DE RECHERCHE

Ces outils concrets, inspirés du *design thinking*, permettent de mettre concrètement en œuvre l'empathie dans les méthodologies de recherche.

3.1. Carte d'empathie

La carte d'empathie est un outil qui permet de se mettre à la place de l'autre et qui aide à synthétiser les actions et à en tirer des insights²³ inattendus. Elle se divise en 4 cases : Dire et Faire – Entendre - Penser – Ressentir. Au fur et à mesure de la lecture des notes d'observation, interviews, ... il faut remplir ces 4 cases.

²³ Un *insight* est une prise de conscience importante d'un fait ou d'une explication.

- **Dire et faire** permet de reprendre les mots clés précis prononcés par l'interlocuteur et de noter les actions et comportements particuliers remarqués pendant les interviews ou observations.
- **Entendre** permet d'appréhender ce que la personne entend de ce que disent les autres, ses amis, ses collègues, les on-dit... qui peuvent l'influencer.
- **Penser** permet de comprendre quelles sont les peurs, les frustrations, les envies et aspirations de la personne.
- **Ressentir** permet de saisir les émotions exprimées par l'individu lors des différentes situations observées.

Souvent, les pensées, sentiments, émotions ne sont pas observés directement mais déduits à partir de divers indices (langage corporel, ton utilisé, tics ou manies, ...). La carte d'empathie a également pour intérêt de mettre en lumière des contradictions, des tensions entre, par exemple, ce que la personne dit et fait. Cela peut être le début de l'identification d'un insight important pour la recherche.

3.2. Cartographie de parcours

Il s'agit au travers de cet outil de créer des diagrammes qui capturent plusieurs informations collectées lors de l'empirie, pour restituer un parcours d'expériences, d'actions ou d'usages. Il faut travailler sur un parcours pertinent pour la problématique (appropriation de nouveaux outils de gestion ou séquence de moments créatifs par exemple), qui permet de capturer comment les faits et actes recueillis se déplacent à travers le temps et l'espace. Il est intéressant d'organiser les données de la recherche d'une manière qui fait sens grâce à une chronologie des événements, voire à des frises parallèles qui permettent de comparer facilement des parcours. Chercher ensuite des récurrences, des faits atypiques, relier des événements individuels à un contexte plus général permet produire des résultats de recherche complets et actionnables.

3.3. Persona

Un *persona* sert à regrouper ses observations clés dans un personnage spécifique, qui constitue une sorte de cobaye pour les réflexions et résultats en cours. Proposé initialement par Cooper en 2004, un *persona* est un archétype représentant un groupe de personnes dont les comportements, motivations et buts sont proches. C'est donc une méthode de synthèse pendant laquelle on crée un personnage semi-imaginaire qui incarne les observations faites sur le terrain (caractéristiques typiques, tendances...). Il est possible de le rendre le plus réel possible en lui donnant un nom, un âge, une anecdote de son enfance, etc. Pour des raisons de manque de temps ou de ressources (financières ou humaines), d'autres types de *personas* peuvent être construits sur des comportements et buts supposés ou imaginés, ou sur des souvenirs. Il s'agira alors de *Provisional personas*, de *Ad-hoc personas* ou de *Proto personas* (Cooper, 2004).

3.4. Check-list

La check-list est utilisée pour déterminer si on est parvenu à un angle d'analyse des données significatif et digne d'intérêt. L'outil original « critical reading checklist » a été développé par D. Larabee (Stanford University). Quatre questions doivent être posées :

1. **A quoi bon ?** Dans quel cadre l'angle d'analyse a été choisi ? Est-il centré sur un acteur ? Un gap théorique ?
2. **Qui le dit ?** Quelle est la validité de l'angle d'analyse choisi ? Cette recherche est-elle corroborée par d'autres recherches ? Est-elle applicable dans un autre contexte ?
3. **Quoi de neuf ?** Quelle est la valeur ajoutée de cet angle d'analyse ? Les conclusions sont-elles articulées d'une nouvelle manière ? Essayer d'être plus précis rendrait-il l'angle d'analyse plus pertinent et nouveau ?
4. **Qui s'en soucie ?** En quoi l'angle d'analyse choisi est important pour la société et vos collègues ? Une reformulation permettrait-elle d'être mieux compris et plus utile ?

3.5. Bodystorming

Lors du travail d'empathie, le *bodystorming* est une méthode unique et utile de recherche d'idées et de prototypage, qui constitue l'expérience physique d'une situation pour en tirer de nouvelles idées. Le *bodystorming* permet de matérialiser les choses, les actes pour mieux les observer et les tester (comme simuler un espace de patients hospitalisés, mettre du gras sur ses lunettes pour voir le monde avec des yeux de personnes déficientes visuelles ou encore simuler une pause-café entre managers). Il faut porter, dans cet exercice, une attention particulière aux prises de décisions liées à l'environnement et aux réactions émotionnelles.

3.6. Prototype

On peut concevoir des artefacts, des situations qui permettent de gagner en empathie, et au final obtenir d'autres informations que celles que l'on récolte habituellement avec des observations ou des entretiens. Un prototype « de l'empathie » permet de mieux comprendre ce que les gens font, disent, pensent et ressentent dans certaines situations.

Quelques idées pour y arriver :

- Demander à votre interlocuteur de dessiner (sa vision du management, sa conception de la créativité dans l'entreprise, ...) et parlez-en avec lui, en lui faisant compléter le dessin au fur et à mesure de vos échanges si nécessaire.
- Créer un jeu de plateau ou de cartes qui aborde les sujets qui doivent être explorés (avec des choix à faire, des priorités à donner, du hasard qui intervient dans les décisions, ...).
- Demander à votre interlocuteur de se mettre dans la peau de quelqu'un concerné par le sujet et de mimer une situation, et d'en parler avec une analogie choisie (animaux, personnage historiques, contes de fées, ...).

3.7. Raconter une histoire

Depuis que le langage existe, les humains ont transmis des informations, ont fait passer des valeurs à travers les contes. Les histoires sont une excellente façon de connecter les gens avec des idées, des concepts. Pour être efficace, l'histoire doit être racontée avec authenticité et avoir trois composantes : une quête (que cherche le personnage ? quelles actions réalisent-

il pour y arriver ?), un conflit (quels obstacles rencontre-t-il ?), une transformation (comment résout-il ce conflit pour atteindre sa quête).

Une histoire bien racontée, avec des détails pertinents, des personnages emblématiques, touche simultanément les auditeurs sur le plan émotionnel et intellectuel.

CONCLUSION

Mobiliser l'empathie dans la recherche en Gestion nous paraît être un enjeu important pour le futur de la recherche si nous voulons comprendre le sens profond des actions des acteurs. De nombreuses autres disciplines ont réfléchi à ce concept, l'ont enrichi, ont proposé des outils de mesure... il est temps que nous fassions de même. Ce chapitre propose des postures et outils, inspirés de la psychologie et du *design thinking*, pour observer et analyser le terrain d'étude avec une approche empathique, mais chacun d'entre nous peut développer ses propres outils pour une parfaite adaptation au contexte de recherche. Ainsi, avec deux collègues, nous avons construit et utilisé un QCM graphique (Dechamp et al., 2015) pour observer les postures d'étudiants à différents moments d'un challenge créatif, en étant le moins intrusif possible (coller des gommettes sur une grande matrice affichée au mur est mieux accepté que répondre à un questionnaire, et on peut, de plus, observer les réactions/émotions associées). Nous avons également, dans cette même recherche photographié tous les matériaux produits par les étudiants (maquettes, représentations graphiques, ...) et demandé aux participants de nous faire part de leur impression sur le processus créatif vécu sous la forme d'un dessin... les résultats sont très parlants !

De même, dans une recherche récente où nous analysons le mode de fonctionnement d'un orchestre sans chef d'orchestre (les Dissonances), nous avons choisi de nous mettre à la place d'un spectateur en suivant les répétitions (avec prise de photo, enregistrements vidéo et prise de notes) et le concert. Au-delà de ce que nous ont dit les musiciens et les spectateurs, nous avons vécu nous-même ces émotions de voir 85 musiciens démarrer une partition, sans chef pour leur indiquer à quel moment le faire. Un moment ... indescriptible ! Nous encourageons donc tous nos collègues chercheurs en Gestion à penser à l'empathie dans leur méthodologie de recherche et à faire preuve de créativité pour développer les outils qui leur seront nécessaires pour y arriver.

RÉFÉRENCES

- Boulanger, C., Lançon, C. (2006). L'empathie : réflexions sur un concept. *Annales médico psychologiques*, 164, 497-505.
- Brown, T. (2008). Design thinking. *Harvard business review*, 86(6), 84.
- Cooper A. (1999). « *The inmates are running the asylum: Why high-tech products drive us crazy and how to restore the sanity*. Sams, 261p.
- Cousins, B. (2018). Design thinking: Organizational learning in VUCA environments. *Academy of Strategic Management Journal*, 17(2),
- Dervin, B., (1992). From the mind's eye of the user : the sense-making qualitative-quantitative methodology (pp. 61-84), in Glazier J. D. et Powell R. R. (éds.), *Qualitative research in information management*, Englewood, CO, Libraries Unlimited, 238 p.

- Dervin, B., (1998). Sense-making theory and practice : an overview of user interests in knowledge seeking and use, in *Journal of Knowledge Management*, 2(2), 36-46.
- Guilmot, N. and Vas, A. (2012). Les cadres intermediaires au cœur du sensemaking: Le processus de création de sens en contexte de changement organisationnel. *Revue internationale de psychosociologie et de gestion des comportements organisationnels*, 18(45), 77-99.
- Hume, D. (1993). *Traité de la nature humaine*. GF Flammarion, Paris.
- Husserl, E. (2003). *La philosophie comme science rigoureuse*. Paris PUF.
- Kotson, I. (2018). *Intelligence émotionnelle et management*. De Boeck supérieur, 4ed.
- Lajante, M. et Lux, G., (2018). Ouvrons la boîte des émotions pour comprendre le processus d'appropriation des outils de gestion ! *Finance, contrôle, stratégie*, 3(21).
- Micheli, P., Wilner, S. J., Bhatti, S. H., Mura, M., & Beverland, M. B. (2019). Doing design thinking: Conceptual review, synthesis, and research agenda. *Journal of Product Innovation Management*, 36(2), 124-148.
- Plattner, H. (2011). *Bootcamp, bootleg: la boîte à outils du design thinking*. Institute of design at Stanford.
- Rogers, C. (1985). Client-centered psychotherapy. In Freedman et al., *Comprehensive textbook of psychiatry*. Baltimore: Williams and Wiley, Wilkins.
- Sharma, G., & Good, D., (2013). The Work of Middle Managers: Sensemaking and Sensegiving for creating Positive Social Change, *Journal of Applied Behavioral Science*, 49 (1), 95-122.
- Tisseron, S., & Bass, H.P., (2011). L'empathie au cœur du jeu social. *Le journal des psychologues*, 3(286), 20-23.
- Weiss, H. M., & Cropanzano, R. (1996). *Affective events theory: A theoretical discussion of the structure, causes and consequences of affective experiences at work*. Elsevier Science/JAI Press.
- Wispé, L. (1987). History of the concept of empathy. In Eisenberg N., Strayer J., *Empathy and its development*. Cambridge University Press.

AUTEUR

Gaëlle Dechamp
 Maître de Conférences HDR en stratégie et entrepreneuriat
 Univ Lyon, UJM-Saint-Etienne, COACTIS, EA 4161, F-42023, Saint-Etienne, France

Mobilizing empathy in our research methods: propositions inspired by design thinking

ABSTRACT

In management sciences, emotions have been of growing interest since the mid-1990s. The role of organizational events as "emotional triggers" have been theorized in affective events theory. The researcher who wishes to understand phenomena and behaviours in and of the company must take this emotional dimension into account and therefore be able to empathize in order to appropriate and understand in depth what is at stake in the organization. The tools of design thinking methodology (a human-centred innovation method, which aims to synthesize analytical and intuitive thinking) are very appropriate to mobilize empathy in management research.

KEY WORDS: Empathy, Emotions, Design thinking, Methodology, Sense-making

INTRODUCTION

Research is informed by multiple experiences, and from their confrontation can emerge a better understanding of the organizational phenomena that we wish to study. In addition, given its scope and variety, it is desirable that management research use a variety of methodologies and paradigms. The researcher must trust himself and sometimes agree to modify his initial analysis plan because the conditions of observation within an organization may change, questioning the initial methodological guidelines. Being "in tune" with the field and its actors, agreeing to make dialectical round trips between objectives, methods and analyses often guarantees the emergence of good research. The researcher's ability to grasp the emotions of the actors he meets, to show empathy in approaching the field and the subtleties of the situations - met and told about - is rarely mentioned in books or articles on research methodologies. However, it is a competence of the researcher that should be valued because, among others, it allows a fine and global analysis of the research field. Of course, psychologists and anthropologists (among others) have developed many methodologies and tools to capture and measure emotions, from which management researchers must draw inspiration. Design thinking, a recent innovation methodology based on a multidisciplinary approach that puts the user, his or her feelings and experiences at the heart of the analysis, has become an interesting source for capturing and understanding emotions.

Based on a systematic review of the literature on design thinking, Micheli *et al.* (2019) proposed nine attributes to define design thinking: abductive reasoning, balance, collaboration, design tools, innovation, problem solving, systemic vision, tolerating failure and user-centricity. These

nine attributes, and especially the tools that accompany them, will lead us to propose another way of approaching research methodology in management sciences by integrating empathy.

1. EMPATHY: A CONCEPT TO BE EXPLORED IN MANAGEMENT RESEARCH

In management research, emotions have been of growing interest since the mid-1990s. The role of organizational events as "emotional triggers" have been theorized in affective events theory (Weiss and Cropanzano, 1996). Since then, several theories of emotion from psychology have been introduced into management science to "show that 1/ organizational behaviour is largely determined by the individual's emotional experience; 2/ the organizational environment is a discrete source of emotional events that trigger these emotions; and 3/ individuals' responses to these events determine the nature of subsequent attitudes and behavioural intentions" (Lajante and Lux, 2018, p.7-8).

The researcher who wishes to understand phenomena and behaviours in and of the company must take this emotional dimension into account and therefore be able to empathize in order to appropriate and understand in depth what is at stake in the organization.

1.1. Definitions

Etymologically (dictionary *le Petit Robert*), the word empathy derives from the Latin radical in- and Im- (en) and the Greek suffix group -pathein (to feel). It expresses the ability to identify with others and to espouse the subjective perception of others.

Empathy is a nomadic concept that is still fragile and defined by several schools of thought. It is distinct from sympathy, which corresponds to emotional contagion. Thus, if empathy is about understanding others, being in sympathy with others is about caring for their well-being (Wispé 1987). Empathy is then "the ability to put oneself in another person's shoes to understand one's feelings, or to represent another's mental representation" (Boulanger and Lançon, 2006). Empathy is expressed through different phenomena such as projection, identification or altruism, and corresponds to a conscious mental simulation of the subjectivity of others.

Empathy is a complex phenomenon, at the heart of the Self, related to emotions and experience (narrative empathy). To understand it, it is necessary to refer to psychological, phenomenological and neuro-scientific theories... but it is Scottish moral philosophy that was the first to introduce the notion of empathy. Boulanger and Sandeel (2006, p. 498) quote Hume who observed in the *Treaty of Human Nature* that "the minds of all men are similar in their feelings and operations, and there is no inclination felt by one man that cannot equally affect all others to some degree." Later, Husserl (1929, 2003) made empathy the key phenomenon on the basis of which intersubjectivity is established as a common behavioural mode. Husserlian empathy is therefore above all a form of cognitive mediation.

Empathy can then be understood not as a mechanism in itself, but as a process of analysis and communication. According to C. Rogers (1985), empathy is about capturing as accurately as possible the internal references and emotional components of another person... and understanding them as if you were that other person.

1.2. A partner to understand the meaning of actions

How about we incorporate empathy into our research methods? Yes, because it is important to understand the meaning of the actors' actions and reflections (also known as "sense-making"). And more particularly for:

- discovering the emotions that guide behaviour
- identifying the right people to talk to
- guiding innovation efforts

The theoretical model of sense-making is presented by Dervin (1992, 1998) as a metaphor characterized by three fundamental concepts: initial situation, lack or discontinuity, help or results. These concepts translate into three questions that indicate the dynamic and iterative nature of the sense-making process: What is missing in your situation and what makes you stop? What questions do you have or what ambiguities do you feel? What kind of help do you expect to get to continue? The individual is called upon to experience different types of discontinuities in a situation taking place in a specific context (Maurel 2010).

Weick, for his part, asks the question: "How can we know what we think (or want) before we have seen (or understood) what we say (or do)?" For him, this sentence contains both the recipe for sense-making and organizing. If meaning makes it possible to organize oneself, it in turn makes meaning possible; one organizes oneself through and for meaning, and the creation of meaning allows organizational actors to share mental models, values and beliefs, to coordinate activities and to act in interrelation. Reality is constructed by the actors (who "enact" their environment) through the interpretations they give it. It is because the creation of meaning is made necessary by the ambiguity inherent in the environment that it is, according to Weick (2001), a central organizational issue.

From a theoretical point of view, Weick (1993) defines the process of meaning creation as a continuous process through which individuals aim to make rational for themselves and others regarding the actions and situations they go through. Sense-making refers to a cognitive process that allows us to grasp the meaning of a situation by assigning it a meaning.

Weick (1993) proposes to approach this process using the ESR (enactment, selection, retention) model. In the enactment phase, the organizational actor tries to give meaning to the context in which he or she finds himself or herself. It seeks to reduce the equivocity of a situation, namely the presence of multiple interpretations for the same situation. The selection phase is about choosing the explanatory scheme acceptable to all from among the different constructed thought schemes. Finally, the retention phase aims to keep the situations encountered, as well as the interpretations made, in the collective memory, so that it can be used when a similar situation arises. The retrospective nature of the sense-making process highlights an organizational learning model. Indeed, it is not possible to give meaning to a fact or action before it has occurred; it is only after experiencing a situation that it is possible to look back to construct its meaning (Argyris & Schön, 2001). Current experiences are then compared with past experiences to create meaning (Guilmot & Vas, 2012, p.84).

The process of constructing meaning is above all specific to each individual since it is marked by cognitive and emotional factors inherent to the human being, whether it is the way he perceives his reality or the way he analyses and interprets elements of information. But this process is also contextual since it takes place in a social, cultural, economic and political environment that characterizes the situation experienced by an individual, group or organization.

The researcher who wishes to better understand the meaning of the actions and reflections of the actors of an organization or to find the emotions that guide behaviors must show empathy. To do this, he or she can rely on methodological proposals, of which many of them are based on design thinking.

2. METHODOLOGICAL PROPOSALS INSPIRED BY DESIGN THINKING

To provide empathy to management research methodologies, it is above all necessary to be in an open and attentive state of mind, and to be ready to shake up traditional methodologies a little to enrich them or exploit them differently. The following four proposals allow the researcher to simply approach the research methodology with a clear resolution of empathy inspired by design thinking. Design thinking is defined as a human-centered method of innovation, and it aims to be a synthesis between analytical and intuitive thinking. It allows a double logic of divergence/convergence (intuition and analogical phases) and a willingness to fully understand experiences (importance of field study, ethnographic approach) (Brown, 2008). Design thinking is generally used in a socio-constructivist approach.

2.1. Adopt the beginner's state of mind

Because we all carry our experiences, our world view and our expertise with us, we have prejudices when we start a study. We all want to succeed in addressing an issue without these prior opinions and let the field tell us what the situation is, but it is not simple. Very precise hypotheses or stereotypes can be false leads that will prove to be obstacles to real understanding of the meaning of the observed persons' actions because they limit empathy. Adopting the state of mind of a novice and putting aside prejudices makes it possible to approach research with new eyes and to avoid missing the meaning given by the actors to their choices and actions. But how can we adopt the novice's state of mind?

1. Try not to judge. Simply observe and interview without orientation, which can lead to value judgments. The way questions are formulated is important in remaining neutral.
2. Question everything... even and especially what you think you already know. Ask questions to learn more about how the person perceives the world. Imagine yourself as a child who wants to understand everything and constantly asks "Why?"
3. Be really curious and try to take a curious posture especially in situations that seem familiar or uncomfortable.
4. Really listen. Do not think about the next topic that will be discussed, but focus on what the interlocutors say, and how they say it.

2.2 Observe more closely with the WHW method (concrete - emotional continuum)

To decode a concrete (what)-emotional (why) continuum in the responses provided or in the observations of actions as broadly as possible, it is interesting to mobilize the WHW (what, how, why) method. These simple questions allow us to move from the concrete observation of a situation to the most abstract emotions and motivations at play in the situation. This is very useful in particular in the analysis of *in-situ* observations, photos or images. We must start with concrete observations (what). What does the person that is observed (or photographed) do? Go into detail, try to be objective and not make assumptions at this stage. Then you have to move on to understanding (how). How does the observed person do what he or she does? Does it require an effort? Does she seem in a hurry, happy, in trouble? It is preferable to use descriptive sentences with precise adjectives. Finally, there is the interpretation phase (why). Why does the person do what they do, and why in this way? This step requires assumptions about the person's motivation and emotions to make sense of the observed situation.

2.3. Prepare and conduct an interview differently

To understand a person's thoughts, emotions and motivations, it is necessary to encourage them to speak richly and completely. The semi-directive interview guides that we use a lot in qualitative management methodology are often too "indicative" and bias responses or close them too early. To interview in an empathetic way, you have to ask why, and again why! Even if you think you know the answer, asking people why they do or say certain things can lead to surprising answers. The conversation arising from a question should continue as long as it seems useful. In addition, the term "usually" should not be used in the questions asked. Instead, it is better to ask people to narrate a specific situation by phrasing the question in the form of "tell me the last time you...". Thus, stories must be encouraged. Whether true or not, they reveal how people think of the world. As for the inconsistencies between what people say and do, they sometimes hide interesting lessons for the researcher. Just like silences, which should not be sought to be filled because they often illustrate a deeper reflection of the person on the current theme. It is also important to avoid closed-ended questions that do not initiate a story, and to avoid questions that might suggest that there is a good answer.

2.4. Synthesize by saturating and then grouping

To complete the empathic approach of a study, it is often relevant to make thoughts, experiences and interviews tangible in the form of keywords or images, through creating short tangible and visual elements conducive to the synthesis and emergence of themes. Grouped on post-it notes, the idea is to display as many as possible (to saturate the wall!) then to group ideas by themes that can guide reflection and analysis. It is then interesting to explore each theme by looking for connections with the other themes or by deepening each idea expressed in the theme. Here we find the same logic as in design thinking ideation sessions, where brainstorming allows us to diverge and then thematic analysis allows us to converge. For greater efficiency, bringing together a few colleagues who are doing the convergence work at the same time as you and/or who note their findings after your groupings is always useful... and is a way to test and discuss a prototype in a way!

Beyond these four methodological postures, precise tools can accompany the researcher in his empathic approach in the field.

3. SUGGESTIONS OF TOOLS TO INCLUDE EMPATHY IN RESEARCH METHODS

These concrete tools, inspired by design thinking, make it possible to concretely implement empathy in research methodologies.

3.1 Empathy map

The empathy map is a tool that puts you in the other's shoes and helps to synthesize actions and draw unexpected insights. The map is composed of 4 boxes: Say and Do - Hear - Think - Feel. As you read the observation notes and interviews, you fill in these 4 boxes.

- Saying and doing makes it possible to perceive the precise key words pronounced by the interlocutor and to note the particular actions and behaviours noticed during interviews or observations.
- Hearing makes it possible to apprehend what the person hears in comparison to what others (his friends, his colleagues, the hearsay or anyone that can influence him or her) say.
- Thinking helps to understand the person's fears, frustrations, desires and aspirations.
- Feeling makes it possible to capture the emotions expressed by the individual during the different situations observed.

Often, thoughts, feelings and emotions are not observed directly but are deduced from various clues (body language, tone used, tics or manias). The empathy map also has the advantage of highlighting contradictions and tensions between, for example, what the person says and does. This can be the beginning of the identification of an important insight for research.

3.2 Course mapping

This tool is a question of creating diagrams that capture several pieces of information collected during the interview to recreate a journey of experiences, actions or uses. It is necessary to work on a relevant path for the problem (the appropriation of new management tools or sequence of creative moments for example), which makes it possible to capture how the facts and acts collected move through time and space. It is interesting to organize the research data in a meaningful way through a chronology of events, or even parallel friezes that make it easy to compare paths. Then searching for recurrences and atypical facts linking individual events to a more general context allows us to produce complete and actionable research results.

3.3. Persona

It is useful to use a persona to group key observations into a specific character, who is created as a kind of guinea pig for ongoing reflections and results. Originally proposed by Cooper in 2004, a persona is an archetype representing a group of people whose behaviours, motivations and goals are similar. It is therefore a method of synthesis during which a semi-imaginary

character who embodies the observations made in the field (typical characteristics, trends...) is created. The character must be made as real as possible by giving him or her a name, an age, a childhood anecdote, etc. For reasons of lack of time or resources (financial or human), other types of personas can be built on supposed or imagined behaviours and goals, or on memories. These will then be Provisional personas, Ad-hoc personas or Proto personas (Cooper, 2004).

3.4. Check-list

The checklist is used to determine whether a meaningful angle of data analysis has been reached. The original "critical reading checklist" tool was developed by D. Larabee (Stanford University). Four questions must be asked:

1. **What's the point?** In which context was the angle of analysis chosen? Is it focused on an actor? A theoretical gap?
2. **Who says so?** What is the validity of the chosen angle of analysis? Is this research supported by other research? Is it applicable in another context?
3. **What's new?** What is the added value of this angle of analysis? Are the conclusions articulated in a new way? Would trying to be more precise make the angle of analysis more relevant and new?
4. **Who cares?** Why is the chosen angle of analysis important for the company and your colleagues? Would reformulation make it possible to be better understood and more useful?

3.5. Bodystorming

This is a unique and useful method for empathy, idea research and prototyping work, which constitutes the physical experience of a situation to draw new ideas from it. Bodystorming makes it possible to materialize things or acts to better observe and test them (such as simulate a space of hospitalized patients, put a barrier on a person's glasses so that he or she sees the world through the eyes of a visually impaired person, or simulate a coffee break between managers). In this exercise, special attention should be paid to environmental decision-making and emotional reactions.

3.6. Prototype

Artifacts can be designed to model situations that allow you to gain empathy and ultimately obtain information other than that that I usually collected through observations or interviews. A prototype of "empathy" helps to better understand what people do, say, think and feel in certain situations.

Some ideas to employ a prototype:

- Ask your interlocutor to draw somethings such as his vision of management or his conception of creativity in the company and discuss it with him, making him complete the drawing as you discuss it if necessary.
- Create a board or card game that addresses the topics that need to be explored (with choices to be made, priorities to be given, chance that comes into play in decisions...).

- Ask your interlocutor to put himself in the shoes of someone concerned by the subject, and mimic a situation or talk about it with a chosen analogy (animals, historical characters, fairy tales...).

3.7. Share a story

Since language has existed, humans have transmitted information and passed values through stories. Stories are a great way to connect people with ideas and concepts. To be effective, the story must be told with authenticity and have three components: a quest (what is the character looking for? what actions does he carry out to achieve it?), a conflict (what obstacles does he encounter?), and a transformation (how does he resolve this conflict to achieve his quest). A well told story, with relevant details and emblematic characters simultaneously touches the listeners emotionally and intellectually.

CONCLUSION

Mobilizing empathy in management research seems to be an important issue for the future of research if we want to understand the deep meaning of the actors' actions. Many other disciplines have thought about this concept, enriched it, and proposed measurement tools... it is time we did the same. This article proposes postures and tools, inspired by psychology and design thinking, to observe and analyze the field of study with an empathic approach. Yet each of us can develop our own tools for a perfect adaptation to the research context. Thus, with two colleagues, we built and used a graphical MCQ (Dechamp et al., 2015) to observe students' postures at different moments of a creative challenge, being as non-intrusive as possible (sticking stickers on a large matrix displayed on the wall is more accepted than answering a questionnaire for example, and we can also observe the associated reactions/emotions). We also, in this same study, photographed all the materials produced by the students (models and graphic representations) and asked the participants to share with us their impressions of the creative process experienced in the form of a drawing... the results are very telling!

Similarly, in a recent study in which we analyze the way an orchestra operates without a conductor (the Dissonances), we chose to put ourselves in the shoes of a spectator by following the rehearsals (with photography, video recordings and note-taking) and the concert. Beyond what the musicians and spectators told us, we personally experienced the emotions of seeing 85 musicians start a score, without a leader to tell them when to do it. That moment was... indescribable! We therefore encourage all our fellow management researchers to think about empathy in their research methodology and to be creative in developing the tools they will need to do so.

RÉFÉRENCES

- Boulanger, C., Lançon, C. (2006). L'empathie : réflexions sur un concept. *Annales médico psychologiques*, 164, 497-505.
- Brown, T. (2008). Design thinking. *Harvard business review*, 86(6), 84.
- Cooper A. (1999). « *The inmates are running the asylum: Why high-tech products drive us crazy and how to restore the sanity*. Sams, 261p.
- Cousins, B. (2018). Design thinking: Organizational learning in VUCA environments. *Academy of Strategic Management Journal*, 17(2),
- Dervin, B., (1992). From the mind's eye of the user: the sense-making qualitative-quantitative methodology (pp. 61-84), in Glazier J. D. et Powell R. R. (éds.), *Qualitative research in information management*, Englewood, CO, Libraries Unlimited, 238 p.
- Dervin, B., (1998). Sense-making theory and practice: an overview of user interests in knowledge seeking and use, in *Journal of Knowledge Management*, 2(2), 36-46.
- Guilmot, N. and Vas, A. (2012). Les cadres intermédiaires au cœur du sensemaking: Le processus de création de sens en contexte de changement organisationnel. *Revue internationale de psychosociologie et de gestion des comportements organisationnels*, 18(45), 77-99.
- Hume, D. (1993). *Traité de la nature humaine*. GF Flammarion, Paris.
- Husserl, E. (2003). *La philosophie comme science rigoureuse*. Paris PUF.
- Kotson, I. (2018). *Intelligence émotionnelle et management*. De Boeck supérieur, 4ed.
- Lajante, M. et Lux, G., (2018). Ouvrons la boîte des émotions pour comprendre le processus d'appropriation des outils de gestion ! *Finance, contrôle, stratégie*, 3(21).
- Micheli, P., Wilner, S. J., Bhatti, S. H., Mura, M., & Beverland, M. B. (2019). Doing design thinking: Conceptual review, synthesis, and research agenda. *Journal of Product Innovation Management*, 36(2), 124-148.
- Plattner, H. (2011). *Bootcamp, bootleg: la boîte à outils du design thinking*. Institute of design at Stanford.
- Rogers, C. (1985). Client-centered psychotherapy. In Freedman et al., *Comprehensive textbook of psychiatry*. Baltimore: Williams and Wiley, Wilkins.
- Sharma, G., & Good, D., (2013). The Work of Middle Managers: Sensemaking and Sensegiving for creating Positive Social Change, *Journal of Applied Behavioral Science*, 49 (1), 95-122.
- Tisseron, S., Bass, H.P., (2011). L'empathie au cœur du jeu social. *Le journal des psychologues*, 3(286), 20-23.
- Weiss, H. M., & Cropanzano, R. (1996). *Affective events theory: A theoretical discussion of the structure, causes and consequences of affective experiences at work*. Elsevier Science/JAI Press.
- Wispé, L. (1987). History of the concept of empathy. In Eisenberg N., Strayer J., *Empathy and its development*. Cambridge University Press.

AUTHOR

Gaëlle Dechamp
Associate professor of strategy and entrepreneurship
Univ Lyon, UJM-Saint-Etienne, COACTIS, EA 4161, F-42023, Saint-Etienne, France

Faire de la recherche-action dans une équipe projet internationale

RESUME

Ce chapitre présente le cas d'une équipe projet mondiale de design et développement, dans laquelle la cheffe d'équipe a mis en œuvre une démarche de recherche-action en deux temps. Durant le travail de l'équipe, elle a adopté une posture de « réflexion en action » permettant de prendre du recul sur les situations et d'adapter le management de l'équipe en conséquence. Puis, après la fin du projet, une collaboration avec une chercheuse n'ayant pas connu l'équipe a donné lieu à une théorisation de l'expérience vécue.

Ce chapitre passe en revue la littérature méthodologique pour montrer en quoi la recherche-action est une méthode intéressante pour la recherche en management international. Puis, il détaille deux situations de l'équipe projet, et la recherche-action menée dans ce contexte. Enfin, des recommandations sont proposées aux chercheurs souhaitant mener une recherche-action en contexte interculturel.

MOTS CLES: Recherche-action, design et développement, recherche de terrain, équipes internationaux, management interculturel.

INTRODUCTION

Le travail en équipes internationales et multiculturelles distantes est en train de devenir la norme dans les entreprises multinationales (Badrinarayanan, 2008). Ces équipes sont souvent chargées de traiter des problèmes complexes et de générer des solutions créatives, et leurs membres sont choisis en fonction de leurs aptitudes et compétences spécifiques (Zander & Butler, 2010). Leur diversité a des conséquences à la fois positives et négatives sur le travail et les performances de ces équipes. La recherche se concentre actuellement sur les contextes et les conditions dans lesquels la diversité culturelle joue un rôle dans les processus et les résultats des équipes (Stahl et al., 2009 ; Zellmer-Bruhn & Gibson, 2014). Il a été montré que la diversité cognitive parmi les membres hétérogènes favorise la créativité, l'innovation et la résolution de problèmes, et se traduit ainsi par des performances supérieures à celles des équipes homogènes (Chua, Roth & Lemoine, 2015; Tadmor et al., 2012).

Cependant, on ne sait toujours pas comment encourager ces résultats positifs dans des équipes interculturelles créatives (Adler & Aycan, 2018). Les recherches sur des sujets connexes en laboratoire ou dans l'enseignement supérieur sont plutôt fréquentes, mais on connaît mal la dynamique des interactions dans ces équipes dans leur contexte opérationnel. La collaboration avec les dirigeants d'équipes multiculturelles au sein d'organisations semble être une opportunité d'accéder au terrain et d'étudier des interactions interculturelles concrètes. L'utilisation de la recherche-action sur le terrain augmente le nombre et les profils d'entreprises au sein desquelles des données peuvent être collectées, offrant ainsi une vue plus complète des organisations et de leurs problèmes de gestion.

Ce chapitre s'appuie sur une recherche-action longitudinale menée par la responsable d'une équipe multiculturelle et multidisciplinaire. L'objectif de ce chapitre est double: d'une part, nous présentons la recherche-action comme une stratégie permettant de générer des informations riches et significatives sur le management d'équipes multiculturelles. La démarche de la recherche-action menée dans cette étude a permis à la cheffe d'équipe (la chercheuse en action) de collecter des données par le biais d'observations d'interactions et de réunions d'équipe, d'enregistrement continu de données, de conversations en profondeur et d'entretiens. Nous décrivons ici notre expérience de « réflexion en action » (Schön, 1983) et comment nous l'avons utilisée pour générer des connaissances théoriques sur la gestion d'équipes multiculturelles. D'autre part, ce chapitre présente quelques événements marquants survenus lors du travail en équipe et explique comment la recherche-action a été utilisée pour corriger le processus de conception et de développement et surmonter les obstacles liés au travail en équipe interculturelle.

Le chapitre est structuré comme suit: premièrement, nous passons en revue la littérature méthodologique pour montrer en quoi la recherche-action est une méthode intéressante pour la recherche en management international. Deuxièmement, nous présentons deux situations découlant du cas dans lequel l'un des auteurs avait mené une recherche-action. Enfin, nous proposons des recommandations aux chercheurs souhaitant appliquer la recherche-action aux contextes interculturels.

1. LA RECHERCHE-ACTION COMME METHODE DE RECHERCHE EN MANAGEMENT INTERNATIONAL

La recherche en management interculturel est souvent réduite à son courant positiviste, historiquement dominant, qui présente les différences culturelles nationales comme un obstacle au travail en équipe interculturelle (Romani, Barmeyer, Primecz & Pilhofer, 2018). Cependant, ce courant de recherche est de plus en plus contesté par les paradigmes de recherche interprétatif, postmoderne et critique (Romani et al., 2018). Ces trois paradigmes reposent principalement sur des méthodologies qualitatives. Le paradigme critique est probablement celui qui est le plus étroitement lié à la recherche-action (Kemmis, 2006). La recherche-action participative a également été utilisée dans le domaine de l'éducation des adultes pour remettre en question les approches positivistes conventionnelles (Zhu, 2019).

La recherche-action contribue à combler le fossé entre la pratique et la recherche, car elle offre un accès privilégié et approfondi au terrain (Gustavsen, 2006; Zhang, Levenson & Crossley, 2015). L'intégration de la théorie et de la pratique, c'est-à-dire le cycle liant expérience pratique et l'application consciente des enseignements tirés de l'expérience, est au cœur de la recherche-action (Schön, 1983). La recherche-action est un processus qui donne du crédit au développement de réflexions, de discussions, de décisions et d'actions par des gens ordinaires participant à une recherche collective sur des « problèmes privés » qu'ils ont en commun (Adelman, 1993). Pour les chercheurs et les praticiens travaillant dans des organisations, la recherche-action offre un moyen d'approcher des problèmes d'organisation par des méthodes scientifiques. De telles méthodes confèrent une validité externe et un soutien à des actions qui pourraient autrement être considérées comme un simple agenda

interne, et biaisées. De plus, les parties prenantes internes engagées dans la recherche-action encouragent le partage d'idées, au profit des praticiens d'autres organisations et de la communauté de la recherche au sens large. Dans ce chapitre, la recherche-action sera définie comme « une approche collaborative de la recherche qui donne aux gens les moyens de prendre des mesures systématiques pour tenter de résoudre des problèmes spécifiques » (Berg, 2004).

Schön (1983) a souligné que le développement professionnel commence par le regard critique qu'une personne sur les événements, et qu'elle met en doute ses actions. Ce doute permet un mode de pensée qui questionne, qui et définit les situations comme des "problèmes". Grâce à une planification minutieuse et à l'élimination systématique d'autres problèmes éventuels, le doute est résolu et les individus sont en mesure de construire leur compréhension de la situation. Ensuite, ils commencent à réfléchir aux situations possibles et à leurs conséquences et à déterminer s'ils ont bien agi.

Il n'existe que quelques exemples de recherches-action menées en situation d'interaction interculturelle. Une école espagnole a utilisé la recherche-action pour catalyser un processus de réflexion pour la transformation de l'école, basé sur une approche interculturelle et inclusive (Sales et al., 2011). Dans le domaine de la gestion, la notion de « réalité négociée » (Friedman & Antal, 2005) s'appuie sur des concepts tirés de la science de l'action et des conflits fondés sur l'identité pour jeter un nouveau regard sur le sens de la compétence interculturelle. Ce cas fait partie des très rares contributions à la littérature.

1.1. Le cas étudié: une équipe internationale de design et développement

Ce chapitre s'appuie sur le cas d'une équipe projet internationale de design et de développement (D&D). L'équipe interdisciplinaire avait pour tâche de développer un meuble technique composé exclusivement de composants originaux, dans l'intention d'un lancement simultané d'un produit identique dans le monde entier deux ans plus tard. L'équipe était composée de quatre sous-équipes situées aux États-Unis, au Royaume-Uni, en Allemagne et en Chine. La première auteure de ce chapitre a participé au projet en tant que cheffe d'équipe et l'a utilisé comme une opportunité de collecte de données, en mobilisant la méthode de réflexion en action (Schön, 1983).

L'équipe D&D globale était composée d'une équipe principale de douze membres et de cinq sous-équipes fonctionnelles (conception, ingénierie, gestion des approvisionnements, opérations, marketing) ou centrées sur des zones géographiques (Asie, Europe, Amérique du Nord et Amérique latine). Par la suite, dans les phases d'innovation en termes de matériaux et de processus de production, des fournisseurs ont également été intégrés à l'équipe de projet "étendue". Ainsi, à mesure que la maturité du projet augmentait et que le délai de mise sur le marché diminuait, la taille de l'équipe projet est passée d'initialement dix membres principaux à plus de 200 membres.

Nous voyons la recherche-action comme centrée sur des connaissances exploitables, et pas nécessairement sur les processus qui transforment les connaissances en actions. Réfléchir sur nos actions menées (« la réflexion en action ») a été un moteur continu tout au long du projet. Cela a été soutenu et encouragé par la culture de l'entreprise, fortement influencée par

les principes indéfectibles de la conception et du design. Mais un tel état d'esprit faisait également partie des exigences de travail des chefs d'équipes D&D de la société. Basée aux États-Unis et possédant des sites en Europe, en Asie et en Amérique latine, la société est mondialement reconnue pour son solide savoir-faire en matière de conception et de développement. Les équipes de D&D mènent souvent plusieurs projets en parallèle. Pour garantir la transmission de l'apprentissage et du savoir-faire des projets antérieurs et d'une équipe de projet à une autre, chaque nouvelle équipe de projet comprend des membres expérimentés, principalement des ingénieurs, mais également des planificateurs de projet et des experts produit.

1.2. Le contexte de la recherche-action

La recherche-action en auto-apprentissage implique un accès constant aux personnes et aux contenus, et agir en tant que praticien réflexif nécessite de nombreuses successions d'essais et d'expérimentation, ainsi que des corrections de cap. Un processus de conception comprend différentes phases de divergence et de convergence, suivies d'une phase de consensus (Munch, 2016). Étant donné que la nature même du processus de recherche-action et des processus de conception et de développement partagent des points communs, le processus de D&D est propice à la recherche par action autonome. En ce qui concerne le cas du projet étudié ici, la recherche a été facilitée par certains facteurs matériels et immatériels.

Les réunions d'équipe et les interactions se déroulaient sur une base hebdomadaire (parfois deux fois par semaine) sur une période de deux ans. Un planificateur de projet à temps plein était chargé de la capture et de l'enregistrement de toutes les données et décisions à des fins de planification et de progression. Les supports ont été proposés sous différentes formes et formats: fichiers texte, images, vidéos, tableurs, croquis, fichiers d'ingénierie et de conception. Ces données ont été téléchargées et organisées dans un référentiel Web derrière le pare-feu de la société, pour consultation par les membres de l'équipe et enregistrement chronologique des événements. Ce référentiel était accessible à tous les membres de l'équipe, quel que soit leur emplacement géographique. Cette richesse de données a été déterminante pour la documentation et la communication des étapes du projet au leadership; il est également devenu le lieu de rencontre où les nouveaux membres de l'équipe qui ont rejoint l'équipe plus tard pouvaient acquérir des connaissances sur le projet. L'accès constant à cette plate-forme a permis la collecte de preuves et la récupération des détails de chaque événement, chaque fois que l'équipe était confrontée à un dilemme majeur.

L'autre facteur qui a facilité le processus de recherche-action était lié à la conception du bureau: l'équipe centrale du projet était située dans un espace ouvert unique, adjacent à une « salle de guerre » du projet. La recherche-action étant une approche collaborative donnant aux personnes les moyens de prendre des mesures systématiques pour tenter de résoudre des problèmes spécifiques (Berg, 2004), la proximité était essentielle. Le processus itératif de vérification, de prototypage, de prise de décision et de révision a été facilité par cette proximité, même pour les membres de l'équipe extérieurs au siège, via de nombreuses sessions de vidéoconférence.

Ce contexte a conduit à deux améliorations remarquables des performances des équipes multiculturelles:

- Le développement de compétences en matière d'échange de connaissances: au fil du temps, les équipes de D & D développent des compétences générales et techniques pour faire face aux problèmes et les résoudre. Équilibrant le risque de réaction exagérée face à de «mauvaises nouvelles» inattendues et le risque de minimiser l'impact de problèmes supplémentaires sur les problèmes en cours de résolution, ces équipes développent des compétences et des processus efficaces d'échange de connaissances.
- L'optimisation des conséquences de la diversité: exemples patents de pluridisciplinarité et de multiculturalisme, les équipes de D & D favorisent l'expression de perspectives diverses au cours de la phase de divergence du processus de conception. Puis, par le biais d'une succession de phases de divergence (génération d'idées) et de convergence (élimination d'idées), ils utilisent des outils et des techniques pour finalement converger vers un objectif commun, la mise sur le marché du produit ou du service.

1.3. Rencontre entre la chercheuse-action et la chercheuse extérieure au terrain

Une fois l'étude longitudinale terminée, la chercheuse-action a confronté son expérience et son analyse à des communautés académiques qui n'avaient pas expérimenté le travail d'équipe. Un universitaire nord-américain lui a suggéré de s'appuyer sur son récit pour le rédiger sous forme d'article et de mener une première analyse de la dynamique multiculturelle en tant que source potentielle d'innovation. Bien que l'hétérogénéité culturelle de l'équipe de projet ait suscité de nombreuses discussions et une diversité de perspectives (comme décrit dans la situation 1), l'auteur ne pouvait toutefois pas mettre en évidence de relations de cause à effet entre la multiculturalité du groupe et le caractère innovant du projet. Cela a certes eu une influence, mais aucune corrélation directe avec le caractère innovant du produit final n'a pu être établie.

La chercheuse a ensuite présenté l'étude de cas lors d'une conférence internationale en management international, où elle a rencontré la seconde auteure de ce chapitre, une universitaire française, experte en management interculturel. Elle a invité la chercheuse-action à approfondir l'analyse du cas à l'aide de nouveaux cadres et perspectives théoriques. En particulier, elle a proposé d'analyser le cas sous l'angle du concept de « systèmes à couplage lâche », qu'elle avait identifié comme une perspective théorique permettant de mieux comprendre le « dilemme consensus / diversité » auquel sont confrontées les équipes multiculturelles (Bachmann, 2006 ; Gibbs, 2006).

Avec l'intégration de cette chercheuse externe à l'étude, un deuxième processus de questions / réflexions a été initié. La chercheuse externe a invité la chercheuse-action à repenser le processus et les événements du travail d'équipe, sur la façon dont les événements se sont déroulés exactement, sur la manière dont ils pourraient être interprétés. La chercheuse externe a suggéré des liens avec des cadres théoriques et des explications sur ce qui s'était passé, ce qui a été réconforté par des retours critiques de la chercheuse-action.

2. ANALYSE DE DEUX SITUATIONS INTERCULTURELLES AU SEIN DE L'ÉQUIPE

Nous allons maintenant présenter et commenter deux leçons tirées de cette étude de cas. Les deux situations illustrent la manière dont l'équipe de projet a été confrontée à des situations interculturelles nécessitant une intervention dans la gestion du projet, de manière proactive et réactive. Certains outils de recherche-action et des approches axées sur la conception ont été mis en œuvre en tant que méthodes supplémentaires pour résoudre les problèmes interculturels lorsque les techniques de gestion traditionnelles étaient insuffisantes pour améliorer les performances de l'équipe. L'originalité de cette approche réside dans l'emprunt d'outils et de techniques de la discipline du design pour relever les défis interculturels en équipe. Nous présenterons d'abord la situation et la solution adoptée pour résoudre le problème, puis nous montrerons l'apport de la recherche-action à deux niveaux, durant le processus d'équipe par la réflexion-en-action, et après la fin de celui-ci par la théorisation ultérieure de la situation vécue.

2.1. Situation 1: l'exercice pre-mortem

En six mois, l'effectif de l'équipe est passé de huit à vingt-cinq membres, et l'hétérogénéité culturelle de l'équipe et le nombre de ses communications asynchrones et virtuelles ont considérablement augmenté. En même temps, l'interdépendance des tâches s'est fait jour et, par conséquent, des problèmes de communication ont apparus: certaines demandes, telles que de prévisions des ventes, sont restées sans réponse ou sont revenues avec des réponses partielles. Incertains du succès des produits, certains responsables de pays en Asie du Sud-Est et au Royaume-Uni ne voulaient pas s'engager sur des chiffres. De même, le nombre de références de produits variait en fonction des marchés et alors que les dirigeants demandaient un lancement progressif (en commençant par un nombre de références restreint), il devenait difficile de parvenir à un consensus sur cette gamme limitée.

Solution adoptée: Un exercice *pré-mortem* impliquant le noyau nord-américain composé de citoyens américains, de la sous-équipe asiatique composée de citoyens de Singapour, de Hong Kong et de la Chine continentale, et de la sous-équipe britannique composée de citoyens britanniques, a été organisé. Les disciplines concernées étaient l'ingénierie, l'ergonomie, la recherche, le marketing et la gestion opérationnelle.

L'exercice exigeait des membres de l'équipe de travailler en petits groupes culturellement mixtes (chaque groupe comprenait au moins un membre de chaque sous-équipe). Six groupes ont été constitués et l'objectif était d'imaginer tout ce qui pourrait mal tourner à un horizon de douze mois après le lancement du produit. En se projetant dans l'avenir et en se mettant à la place des autres membres de l'équipe ou des parties prenantes (clients, partenaires, collègues), les membres de l'équipe étaient invités à faire part de leurs préoccupations en petits groupes et à faire preuve d'empathie pour les personnes potentiellement touchées par ces problèmes. En plus de cet «exercice dystopique», chaque petite équipe a eu le temps de travailler ensemble, de régler les problèmes et de présenter à l'équipe élargie ses recommandations appuyées sur un ensemble de solutions permettant de réduire les risques. Quand les problèmes ne pouvaient pas être résolus par l'équipe qui les avait exprimés, l'aide de toute l'équipe était requise. Toutes les solutions ont été enregistrées. La liste de problèmes potentiels avec les solutions correspondantes a été imprimée sur de grandes affiches qui sont

restées pendant toute la durée du projet affichées sur les murs de la « salle de guerre ». Au cours du projet, il a souvent été fait référence à ces solutions envisagées, car certaines problèmes sont devenus une réalité. En résumé, tous les membres de l'équipe avaient interagi de manière étroite, avaient réduit leur niveau d'anxiété et avaient généré des idées dans un contexte peu contraignant : ils étaient ainsi mieux préparés pour envisager un large éventail de solutions, le cas échéant.

Réflexion-en-action : Dans le processus de conception, la cheffe d'équipe a inclus une session qui n'était pas prévue à l'origine, dans le but de déclencher l'expression de craintes personnelles ou liées au projet. Comme décrit ci-dessus, l'objectif était d'identifier les difficultés interculturelles potentielles avant qu'elles ne surviennent. En formant de petits groupes à forte diversité culturelle et fonctionnelle, un contexte favorable à l'expression des préoccupations et des peurs a été créé. Chacun des six groupes a enregistré toute une liste de problèmes professionnels et personnels. Afin de respecter les préférences du groupe en matière de travail et de surmonter les mauvaises interprétations linguistiques, les membres de l'équipe ont été invités à réfléchir individuellement et à rédiger leurs points de vue par écrit avant de se regrouper. Certaines équipes ont choisi de partager et de discuter des problèmes au cours de deux ou trois séances.

Le processus était un processus collaboratif d'échange de connaissances, dans un climat de diversité axé sur l'engagement. Les petites équipes se sont rapidement approprié la situation et les problèmes. Cet exercice était de nature spéculative qui visait à réduire l'impact des conflits à venir, ainsi que les craintes, a en fait accru les liens de coopération entre les membres de l'équipe. Les collègues nord-américains ont offert à leurs collègues d'Asie et du Royaume-Uni une occasion d'exposer leurs points de vue en dehors de toute pression hiérarchique ou de temps. L'exercice a également révélé la nature des craintes et des préoccupations de chacun, sans avoir à traverser les difficultés émotionnelles liées à un problème réel. Comme la composition et les avis des équipes étaient affichés sur les murs de la salle de projet, il était possible d'identifier plus tard les personnes les moins à l'aise avec un événement ou une situation et d'en discuter de manière rationnelle. Ces informations ont aussi permis à la cheffe d'équipe d'aller chercher de manière proactive du soutien ou des ressources supplémentaires pour résoudre les problèmes.

Théorisation: La collaboration avec une chercheuse externe (la deuxième auteure) a permis d'analyser cette situation à travers le prisme de la théorie des systèmes. Si on envisage l'équipe de design et développement sous l'angle des systèmes à couplage lâche (Bachmann, 2006 ; Orton & Weick, 1990), l'exercice pré-mortem fut axé sur la tâche à accomplir (couplage concernant les aspects structurels) par le biais d'une interaction sociale et interculturelle (couplage concernant les aspects culturels et sociaux). Les membres de l'équipe n'étaient pas obligés de s'entendre sur des valeurs communes ou adopter des comportements particuliers (pas d'obligation de couplage culturel serré), mais plutôt d'adopter une perspective empathique, qui a finalement abouti à un resserrement du couplage culturel.

À travers le prisme du design et selon l'interprétation du chercheur en action, le designer se positionne souvent comme un modérateur, orchestrant les aptitudes et les activités nécessaires pour assurer une participation démocratique et productive (Light & Akema, 2012). Des études récentes axées sur les facteurs de succès des projets de co-design révèlent un

nouveau rôle pour les designers en tant que « générateurs de design », opérant dans les registres de la collaboration pour mixer experts en abstraction et personnes pragmatiques afin de transformer le mode de travail des professionnels en insérant une pensée design plus créative (Lee, 2018). Cette compréhension conduit à l'interprétation suivante: alterner les phases de divergence dans la recherche de connaissances ou d'idées puis les phases de convergence pour sélectionner et tester les idées les plus prometteuses, peut représenter un défi pour les managers d'équipes interculturelles, mais un défi très courant en design. Le designer, à force d'adapter constamment son niveau de communication entre experts en abstraction et personnes pragmatiques, développe une capacité d'écoute et de compréhension de toutes les nuances au sein de son équipe, au profit des différents acteurs dont les vues et idées sont intégrées dans la formulation de solutions possibles.

2.2. Situation 2: le positionnement-prix en Asie

L'équipe marketing d'Asie basée à Singapour était préoccupée par un positionnement prix trop élevé, car elle n'avait pas réussi à convaincre le siège et les gestionnaires du marché chinois à s'entendre sur la différence de prix entre les États-Unis et les marchés asiatiques. Par conséquent, l'organisation commerciale en Asie n'était pas prête à s'engager sur ses prévisions de ventes, ce qui devenait très problématique pour évaluer la viabilité financière du projet et les quantités de pièces prévisionnelles par fournisseur. Plusieurs réunions infructueuses ont eu lieu entre l'équipe centrale américaine, les sous-équipes d'Asie et du Royaume-Uni, les ingénieurs britanniques et américains, ainsi que les responsables des approvisionnements britanniques et américains.

Solution adoptée : La responsable marketing pour l'Asie, basée à Singapour, a envoyé à l'équipe centrale un document détaillé exposant ses préoccupations, et a demandé qu'une discussion soit organisée dès réception du document. Le document était bien rédigé et documenté. Certaines attentes étaient prévisibles, comme un positionnement prix plus bas, un nombre réduit d'options, ou encore le souhait d'une palette de couleurs saturée. Mais d'autres éléments ont surpris l'équipe centrale: les marchés asiatiques demandaient un nombre élevé de « composants fabriqués localement » pour répondre à des objectifs de développement durable, ainsi qu'un plan de lancement ambitieux. Le document a suscité de nombreuses discussions au sein de l'équipe centrale.

Réflexion-en-action : La cheffe de l'équipe centrale a compris que la crédibilité de l'équipe asiatique était en jeu et a décidé de faire un pas en leur direction. L'équipe centrale a annoncé officiellement que les coûts nécessaires pour permettre un positionnement prix aussi bas en Asie ne pouvaient être atteints, et que les écarts de prix entre les différentes régions du monde ne pouvaient dépasser un certain pourcentage. Quelques jours plus tard, l'information était officiellement partagée avec les gestionnaires des marchés asiatiques. Le directeur général pour l'Asie demanda une conversation directe avec la cheffe de l'équipe centrale, qui déclencha une analyse complète des coûts avec les dirigeants de toutes les autres régions d'Europe et des Amériques. Cette initiative était le point de départ d'une importante mission d'approvisionnement global pour le responsable des approvisionnements de l'équipe centrale, qui aboutit finalement à une réduction significative des coûts de production locaux.

Théorisation : Pour reprendre le cadre d'analyse de l'équipe interculturelle en tant que système à couplage souple (Bachmann, 2006 ; Orton & Weick, 1990), la sous-équipe asiatique a renforcé le couplage structurel (en rédigeant par écrit sa vision du produit pour le marché asiatique), ce qui a provoqué un couplage culturel renforcé (meilleure compréhension de la vision du monde de l'autre par la conversation directe avec la cheffe d'équipe). Ensuite, cela a provoqué de nouveau un couplage structurel plus serré (via la mission d'approvisionnement).

Les préoccupations des entreprises liées aux spécificités culturelles, en particulier en ce qui concerne le positionnement prix sur les marchés locaux, sont des sources fréquentes de conflits interculturels qui conduisent souvent à des situations d'impasse (aucune partie ne veut céder). Créer un lieu de conversation sûr pour l'équipe asiatique a permis de débloquent d'autres problèmes jusqu'alors inconnus de l'équipe. De plus, en invitant les gens à échanger et à exprimer leurs préoccupations sous une forme écrite qui pouvait être partagée avec les dirigeants, l'équilibre dans les relations de pouvoir entre l'Asie et l'Amérique du Nord a été rétabli. Cette ouverture s'est faite grâce à la combinaison de deux méthodes: 1. la réflexion en action qui a offert un nouveau cadre de partage des préoccupations et 2. le management interculturel qui a créé un contexte permettant d'exprimer en toute sécurité des exigences fortes et totalement inattendues de la part de l'autre culture.

3. QUELS CONSEILS POUVONS-NOUS DONNER AUX CHERCHEURS QUI ENVISAGENT DE FAIRE DE LA RECHERCHE-ACTION EN CONTEXTE INTERCULTUREL?

3.1. La recherche-action peut-elle être menée dans toute équipe ou organisation?

La portée et l'orientation de la réflexion du dirigeant en action sont fortement influencées, et peuvent être sévèrement limitées, par les systèmes d'apprentissage de l'organisation dans laquelle il exerce (Schön, 1983). Par conséquent, la recherche-action n'est probablement pas appropriée dans tous les contextes organisationnels. La culture organisationnelle doit être suffisamment ouverte pour permettre la succession d'essais et d'erreurs, et suffisamment ouverte dans ses valeurs et ses convictions pour permettre aux propositions issues de la détection et la correction des erreurs d'enrichir l'apprentissage collectif (Argyris & Schön, 1978). Ce cas met en évidence la recherche-action au sein d'une équipe D&D, où essais et erreurs font partie des activités quotidiennes. Les dirigeants de D&D sont invités à intégrer dans leurs activités la reconnaissance et la correction d'une faute ou d'une erreur perçue (Argyris & Schön, 1978), pour qu'en tant que porteurs du modèle de « pensée lean » de Toyota, ils puissent aussi éliminer le « gaspillage » dans les environnements tertiaires (Womack & Jones, 1996). Si la culture organisationnelle le permet et que le chef d'équipe s'intéresse à la pratique de la recherche-action, il n'y a aucune raison pour que cela ne soit pas essayé.

3.2. Tout acteur d'une organisation peut-il faire de la recherche-action?

Au cours des différentes étapes d'un nouveau projet, les équipes doivent souvent prendre des décisions en l'absence d'informations préalables adéquates - ou de règles - pour appuyer ces décisions. Certaines entreprises croient aux tests de marché, aux groupes de discussion, etc. D'autres s'appuieront sur leurs propres expériences de recherche centrées sur l'utilisateur et le dynamisme intuitif de leurs créatifs.

La recherche-action requiert des compétences préalables particulières, qui peuvent être apprises par l'acteur de l'organisation qui souhaite faire de la recherche-action: des compétences d'écoute, de rigueur et de discipline, le désir inlassable de trouver la « vérité » (ou les effets des symptômes identifiés), et une croyance selon laquelle « il y a toujours une meilleure façon de faire ». Du point de vue personnel, les caractéristiques requises sont la curiosité, l'empathie, la générosité (au sein d'un leadership partagé), la transparence (la communication ouverte est primordiale), le questionnement critique et la capacité de toujours prendre position lorsqu'il s'agit d'avancer.

Dans le contexte du management interculturel, les deux situations décrites ci-dessus décrivent certaines initiatives spécifiques qui ont permis de surmonter les difficultés. La première situation met en exergue une approche proactive de la gestion interculturelle: elle offre aux membres de l'équipe une marge de manœuvre pour envisager l'avenir et discuter en petits groupes de « ce qui pourrait mal tourner ». Ceci leur donne un terrain d'expression pour leurs craintes et préoccupations, par écrit et en face-à-face. Les petits groupes ont créé un contexte de sécurité et de confiance, qui a apporté à la cheffe d'équipe une richesse de perspectives culturelles et des évaluations de la viabilité du projet sous divers angles.

La deuxième situation met en évidence le fait que l'équipe principale a reconnu qu'une sous-équipe avait besoin de l'appui d'un groupe plus important pour transmettre un message difficile à sa hiérarchie locale. Être à l'écoute constante de signaux subtils est une compétence qui doit être présente et constamment développée chez le « chercheur en action ».

3.3. Quelle méthodologie est appropriée pour la recherche-action?

S'ils entendent approfondir l'analyse des données et la publication de leurs recherches, les chercheurs en action doivent archiver un maximum de données empiriques, notamment des données « produites par une action organisationnelle » (rapports de décisions, prototypes, comptabilité et autres chiffres, etc.), ainsi que des données de processus. Celles-ci comprennent des versions successives de différents documents, mais également des notes sur les interactions (qui ? pourquoi ? quoi? etc). Globalement, il en résultera probablement un nombre considérable de documents et de données qu'il sera difficile de trier et d'analyser ultérieurement. Par conséquent, la tenue d'un « journal de recherche » est fortement recommandée aux chercheurs en action afin qu'ils puissent enregistrer leurs événements et les dates critiques. Le journal de recherche est utile sous trois aspects: 1) il fournit une « version courte » des événements et processus étudiés, 2) il aide à récupérer les données critiques dans la masse de données produites (par exemple, triées par date d'occurrence), et 3) il déclenche des points de réflexion sur le cas étudié.

Comme décrit dans ce chapitre, une réflexion quotidienne est essentielle. Selon les cas, cela peut se produire en fin de journée ou le lendemain matin, lors de la relecture d'événements ou d'activités passés. Une autre possibilité est d'impliquer un groupe élargi de contributeurs dans ce processus de prise de recul, afin de réfléchir à des problèmes plus complexes.

3.4. Est-ce utile de collaborer avec des chercheurs externes?

La collaboration avec des chercheurs externes est importante pour la recherche-action. Le chercheur en action expérimente le processus empirique « de l'intérieur », et un chercheur externe examine ce processus du point de vue d'un tiers. La combinaison des deux est très bénéfique. Le chercheur en action contribue par ses connaissances empiriques approfondies et sa compréhension experte de la situation, mais celles-ci peuvent être révélées et poussées plus loin lorsqu'elles sont mises au défi par un chercheur externe.

La contribution d'un chercheur externe est essentielle pour fournir une interprétation appuyée sur des cadres théoriques. Comme décrit précédemment, les praticiens qui s'engagent dans la recherche-action réagissent souvent de manière intuitive aux situations, et la théorisation profite grandement de l'afflux de théories et de concepts testés dans d'autres domaines ou contextes. Les praticiens observent et ressentent les mêmes phénomènes que les chercheurs mais ne savent parfois pas comment les nommer, ni comment les relier aux cadres théoriques existants.

Idéalement, le dialogue entre le chercheur en action et le chercheur externe devrait être mené parallèlement au processus empirique (Bartunek, Avenier & Bloch Dolande, 2015). Si cela n'a pas pu se produire, il peut toujours être utile de discuter de la situation ultérieurement, comme cela fut le cas ici.

CONCLUSION

Ce chapitre s'est concentré sur l'application de la méthodologie de recherche-action à un processus de management interculturel. Deux situations extraites d'une étude de cas ont montré comment un praticien avait utilisé la réflexion en action pour apporter des réponses innovantes à des situations problématiques qui auraient pu conduire à une impasse ou retarder considérablement le projet. Dans les deux cas, le défi consistait à amener les personnes de différentes cultures à exprimer en toute sécurité leurs préoccupations, leurs appréhensions ou leurs demandes. Il est souvent difficile de créer un contexte sûr pour prendre la parole, en particulier vis-à-vis du groupe culturel qui détient le leadership en termes de catégorisation sociale ou de langage (Zhang & Guttormsen, 2015). La cheffe d'équipe a servi de passerelle aux membres de l'équipe issus de divers contextes culturels pour accéder aux activités quotidiennes du projet, et leur a permis de le faire de manière transparente et hautement interactive. Ce chapitre montre que l'approche de réflexion en action, combinée à la compétence interculturelle, a permis de débloquer deux situations qui auraient pu aggraver la communication au sein de l'équipe (peur et angoisse de l'échec du projet) ou créer un conflit critique entre les équipes nord-américaine et asiatique. La résolution du « dilemme diversité /

consensus » s'est faite sans compromettre la diversité des mentalités en suscitant un lieu sûr pour prendre la parole et défendre ses idées. En même temps, un débat factuel intégrant les diverses attentes du marché, sans compromettre le produit, a permis de dégager un consensus. En fait, cela a amélioré le produit et l'a rendu plus respectueux de l'environnement en Asie que prévu initialement.

Après avoir examiné la méthodologie adoptée et l'avoir comparée aux cadres existants en management international, nous pensons que la recherche-action peut être menée par la plupart des leaders d'équipes interculturelles, mais pas nécessairement dans toutes les organisations. Le contexte organisationnel doit être fondé sur la liberté d'exploration et d'expression, la pensée critique et des efforts incessants portés vers l'apprentissage et l'amélioration continue.

BIBLIOGRAPHIE

- Adelman, C. (1993). Kurt Lewin and the Origins of Action Research. *Educational Action Research*, 1:1, (7-24).
- Adler N. J., Aycan Z. (2018). Cross-cultural Interaction: What We Know and What We Need To Know. *The Annual Review of Organizational Psychology and Organizational Behavior*, 5:307-33.
- Argyris, C., & Schön, D. (1978) *Organizational learning: A theory of action perspective*, Reading, Mass: Addison Wesley.
- Avenier, M. J., & Schmitt, C. (2009). Un cadre méthodologique pour des recherches tirant parti de l'expérience de praticiens de la gestion en PME. *Economies et Sociétés – série Economie de l'entreprise*, 43(2), 271–294.
- Bachmann, A. S. (2006). Melting Pot or Tossed Salad? Implications for Designing Effective Multicultural Workgroups from a Coupling Perspective. *Management International Review*, 46 (6): 721–747.
- Badrinarayanan, V. (2008). Effective virtual new product development teams: An integrated framework. *Journal of Business & Industrial Marketing*, 23: 242–248.
- Bartunek, J. M., Avenier, M. J., & Bloch Dolande, M. (2015). Dialogics and Academic-Practitioner Dialogues. In *Academy of Management Proceedings 2015* (1), 127-98. Briarcliff Manor, NY 10510: Academy of Management.
- Berg, B. (2004). *Qualitative research methods for the social sciences*, Boston. Pearson Education, 5th edition, Ch.7.195–207.
- Chua, R. Y., Roth, Y., Lemoine, J. F. (2015). The impact of culture on creativity: How cultural tightness and cultural distance affect global innovation crowdsourcing work. *Administrative Science Quarterly*, 60(2): 189-227.
- Friedman V., Antal, A.B. (2005). Negotiating Reality: a Theory of Action Approach to Intercultural Competence. *Management Learning*, 36 (1), 69–86.
- Gibbs, J. L. (2006). Decoupling and coupling in global teams: implications for human resource management. In: G. K. Stahl and I. Björkman. *Handbook of Research in International Human Resource Management*. Cheltenham, Edward Elgar Publishing: 347–363.
- Gully, S. M., Devine, D. J., & Whitney, D. J. (1995). A meta-analysis of cohesion and performance: Effects of level of analysis and task interdependence. *Small group research*, 26 (4), 497–520.

- Gustavsen, B. (2006). Theory and Practice: the Mediating Discourse. In: P. Reason and Hilary Bradbury: *Handbook of Action Research*. London: SAGE: 17-26.
- Hajro, A., Gibson, C., & Pudelko, M. (2017). Knowledge Exchange Processes in Multicultural Teams: Linking Organizational Diversity Climates to Teams' Effectiveness. *Academy of Management Journal*, 60(1), 345–372.
- Kemmis, S. (2006). Exploring the Relevance of Critical Theory for Action Research: Emancipatory Action Research in the Footsteps of Jürgen Habermas. In: P. Reason and Hilary Bradbury: *Handbook of Action Research*. London: SAGE: 94-105.
- Lee, Y. (2008). Design participation tactics: the challenges and new roles for designers in the codesign process. *International Journal of Co-Creation in Design and the Arts*, 4 (1), 31–50.
- Lewin, K. (1946). Action research and minority problems. *Journal of Soc. Issues* 2(4): 34–46
- Light A., Akama, Y. (2012). The human touch: from method to participatory practice in facilitating design with communities. In *Proceedings of the Participatory design conference: PDC2012*. Roskilde, Denmark: ACM.
- Münch, F. (2016). From Europe to North America to Asia: Overcoming the Hurdles of Interdisciplinary Multicultural Teams through a Design-Driven Process. *Management International/International Management/Gestión Internacional*, 20(Special Issue): 38-48.
- Orton, J.D., K.E. Weick (1990). Loosely coupled systems. A reconceptualization. *Academy of Management Review*, 15(2): 203-223.
- Romani, L., Barmeyer, C., Primecz, H., & Pilhofer, K. (2018). Cross-cultural management studies: state of the field in the four research paradigms. *International Studies of Management & Organization*, 48(3), 247-263.
- Romme, A. G. L., Avenier, M. J., Denyer, D., Hodgkinson, G. P., Pandza, K., Starkey, K., & Worren, N. (2015). Towards common ground and trading zones in management research and practice. *British Journal of Management*, 26 (3), 544–559.
- Sales, A., Traver J.A., Garcia R. (2011). Action research as a school-based strategy in intercultural professional development for teachers. *Teacher and Teaching Education*, 27 (5), 911–91.
- Schön, D. (1983). The Reflective Practitioner. *How Professionals Think in Action*. New York, NY: Basic Books.
- Stahl, G.K., Maznevski, M.L., Voigt, A., Jonsen, K. (2009). Unraveling the effects of cultural diversity in teams: A meta-analysis of research on multicultural work groups. *Journal of International Business Studies*, 41(4): 690-709.
- Tadmor, C. T., Satterstrom, P., Jang, S., Polzer, J. (2012). Beyond Individual Creativity: The Superadditive Benefits of Multicultural Experience for Collective Creativity in Culturally Diverse Teams. *Journal of Cross-Cultural Psychology*, 43(3), 384-392.
- Womack J.P., & Jones, D.T. (1996). *Lean Thinking*, New York. Simon & Schuster.
- Zander L., Butler, C.L., (2010). Leadership modes: Success strategies for multicultural teams. *Scandinavian Journal of Management*, 26: 258-267.
- Zellmer-Bruhn ME, Gibson C. (2014). How does culture matter? A contextual view of intercultural interaction in groups. In M. Yuki, M. Brewer (eds.): *Culture and Group Processes*, New York: Oxford University Press: 342-402.
- Zhang W., Levenson A, & Crossley C., (2015). Move your Research from the Ivy Tower to the Board Room, *Human Resource Management*, 54 (1), 151–174
- Zhang, L.E, Guttormsen, D.S.A (2015). Multiculturality as a key methodological challenge during in-depth interviewing in international business research. *Cross Cultural & Strategic Management*, 23(2), 232–256.

Zhu, Y. (2019). Can participatory action research empower participants in adult education studies ? *The Canadian Journal for the Study of Adult Education*, 31(1), 63-73.

AUTEURS

Fabienne Münch

Executive Director, Global Initiatives and Strategy; University of Chicago, USA
Doctorante en Management International
Univ. Grenoble Alpes, Grenoble INP, CERAG, F-38000 Grenoble, France

Anne Bartel-Radic

Professeur des universités en Management International
Univ. Grenoble Alpes, Grenoble INP, CERAG, F-38000 Grenoble, France

Doing action research in a global project team

ABSTRACT

This chapter presents the case of a global design and development project team in which the team leader has implemented two phases of action-research. During the teamwork, she implemented "reflection in action" that allowed to take a step back on situations and adapt the management of the team accordingly. Later, after the end of the project, a collaboration with an external researcher led to further theorization of the teamwork experience.

This chapter first reviews the methodological literature to show that action research is an interesting method for international management research. Then, it presents two events that occurred during the project teamwork, and how action research was conducted in this context. Finally, recommendations are made to researchers who intend to implement action research in intercultural teams.

KEY WORDS: Action research, design and development, field research, global teams, intercultural management, research in action.

INTRODUCTION

Virtual teamwork or distributed multinational and multicultural teams are becoming the norm in today's business world for global team operations (Badrinarayanan, 2008). Teams are frequently formed to address multifaceted issues, solve complex problems, and generate creative solutions, as members are generally chosen based on their specific skills and competencies (Zander & Butler, 2010). Team diversity has both positive and negative consequences on teamwork and team performance, and contemporary research focuses on the contexts and conditions under which cultural diversity plays a role in team processes and outcomes (Stahl et al., 2009, Zellmer-Bruhn & Gibson, 2014). Cognitive diversity among heterogeneous members has been shown to promote creativity, innovation and problem solving, and thus results in superior performance than homogeneous teams (Chua, Roth & Lemoine, 2015; Tadmor et al., 2012).

However, it is still insufficiently known how positive outcomes can be more accurately perceived, and encouraged, in culturally diverse creative teams (Adler & Aycan, 2018). While research on related topics abound in the lab or in higher education, understanding the dynamics of team interactions in their operational context is lacking. Collaboration with leaders of multicultural teams in organizations appears to be an opportunity to access the field and study hands-on intercultural experiences. The use of a field-based, action research approach, expands the number and profiles of organizations from which data can be collected, providing a more comprehensive view of organizations and their management issues.

This chapter draws on a longitudinal action research led by the leader of a multicultural and multidisciplinary team. The objective of this chapter is twofold: on the one hand, we present action-research as a strategy for generating rich and meaningful insights on multicultural team management. The self-led action-research approach used in this study enabled the team leader (the action-researcher) to collect data through observations, team meeting interactions, continuous data recording, in-depth conversations and interviews. Here, we relate our experience of reflection-in-action (Schön, 1983), and how we used it to generate theoretical insights on multicultural team management. On the other hand, this chapter presents some salient events that occurred during teamwork, and how action-research was used to course-correct the design and development process and overcome hurdles of intercultural teamwork. The chapter is structured as follows: first, we review methodology literature and argue that action research is an interesting method for international management. Second, we present two situations originating from the case over which action research had been led by one of the authors. Third and finally, we offer recommendations for scholars interested in applying action research to intercultural contexts.

1. ACTION RESEARCH AS A RESEARCH METHOD FOR INTERNATIONAL MANAGEMENT

Cross-cultural management research is often reduced to its historically dominating, positivist stream, which is presenting national cultural differences as hurdles for intercultural teamwork (Romani, Barmeyer, Primecz & Pilhofer, 2018). However, this stream of research is increasingly challenged by the interpretivist, the postmodern, and the critical research paradigms (Romani et al., 2018). These three paradigms mainly rely on qualitative methodologies. The critical paradigm is probably the one that is the most closely related to action research (Kemmis, 2006). Participatory action research has also been used in the area of adult education to challenge conventional positivist approaches (Zhu, 2019).

Action research participates in bridging the divide between practice and research, because it offers privileged and in-depth access to the field (Gustavsen, 2006; Zhang, Levenson & Crossley, 2015). Central to action research is the integration of theory and practice, which means the cyclic pattern of experience and the conscious application of lessons learned from experience (Schön, 1983). Action research is a process that gives credence to the development of reflective thoughts, discussion, decision and action by ordinary people participating in collective research on “private troubles” that they have in common (Adelman, 1993). For researchers and practitioners working in organizations, action research offers a way to scientifically evaluate important organizational issues by applying scientific methods. Such methods provide external validity and support for actions that otherwise might be dismissed as representing a biased internal agenda. Additionally, internal stakeholders engaged in action research foster the sharing of insights, to the benefit of other organizations’ practitioners and the broader research community. In this chapter, action research will be defined as “a collaborative approach to research that provides people with the means to take systematic action in an effort to resolve specific problems” (Berg, 2004).

Schön (1983) stressed that professional growth really begins when a person starts to view things with a critical lens by doubting about his or her actions. Doubt brings about a way of

thinking that questions and frames situations as "problems". Through careful planning and systematic elimination of other possible problems, doubt is settled, and individuals are able to affirm their knowledge of the situation. Then, they start to think about possible situations and their subsequent outcomes and consider whether they carried out the right actions.

Only few examples of connecting action research and intercultural interaction exist in literature. A Spanish school used action research to serve as a catalyst for a reflective process for school transformation, based on an intercultural and inclusive approach (Sales et al., 2011). In the field of management, the concept of "negotiating reality" (Friedman & Antal, 2005) draws on concepts from action science and identity-based conflict, to take a new look at the meaning of competence in intercultural interaction. This case is among the very few contributions made to the literature.

1.1. The case study: a global design and development team

The chapter sheds light on the creation of a global design project by an international Design and Development (D&D) team. The interdisciplinary team's task was to develop a global product made exclusively of original components, with the intent to launch the exact same product worldwide two years later. The team was composed of four sub-teams located in the United States, United Kingdom, Germany and China. One of the authors participated in the project as the team leader and used the project as a field for data collection through the method of reflection-in-action (Schön, 1983).

The D&D team structure consisted of a core team of twelve members, and five sub-teams that were either disciplinary driven (design, engineering, supply management, operations, marketing), or geographically focused (Asia, Europe, North America and Latin America). Later on, when the project called for materials and production process innovation, vendors became part of an "extended" project team. Thus, as project maturity raised and time to market shrank, the project team that started with ten original core members became as large as 200 members. Our view of action research is that it more generally focuses on actionable knowledge, and not necessarily on the processes that turn knowledge into action. Reflecting on our actions as we were doing them ("reflection-in-action") was a continuous driver during the course of the project. In part, it is to be attributed to the company culture, heavily influenced by unyielding tenets of design and design-driven process in major corporate decisions: this culture spans across the recruitment of new office workers, assembly lines implementation principles, or in rolling out its marketing strategies. But such a mindset was also part of the company D&D team leaders' work requirements. Headquartered in the United States with offices in Europe, Asia and Latin America, the company is worldly recognized for its strong design & development know-how; several D&D teams lead multiple projects in parallel. To ensure that past projects' learning and know-how transfer happen from one project team to another, every new project team includes senior project members, mainly engineers, but also project planners and product experts.

1.2. Action research context

Self-study action research implies constant access to people and content and acting as a reflective practitioner requires numerous successions of trials and experimentation, and course correction. A design process includes alternate phases of divergence and convergence, followed by a phase of consensus (Munch, 2016). Since the very nature of the action research process and the design and development process share commonalities, the D&D process is conducive to self-action research. As far as the case considered, the research was eased by a few factors, some material, some immaterial.

Team meetings and in-person interactions happened on a weekly basis (sometimes twice a week) over a period of two years. A full-time project planner was in charge of capturing and recording all data and decisions for planning and progress purposes. Support materials came in various shapes and format: text files, pictures, videos, spreadsheets, sketches, engineering and design files. These data were uploaded to, and organized under, a web-based repository behind the company firewall, for team members' consultation and chronological recording of events. That repository was accessible to all team members, regardless of their geographic location. This wealth of data has been instrumental for the documentation and reporting of project stages to leadership; it also became the "go to" place where new team members who joined the team down the road could garner prior project knowledge. Constant access to this platform enabled the collection of evidence, and retrieval of any given event details, whenever the team faced a major dilemma.

The other factor that eased the action research process was linked to office design: the project core team was located within a single open space, adjacent to a project "war room". As action research is a collaborative approach that provides people with the means to take systematic action in an effort to resolve specific problems (Berg, 2004), proximity was key. The iterative process of checking, prototyping, decision making, revisiting, was eased by this proximity, even for team members outside of the headquarters, via numerous videoconferencing sessions.

This context leads to two noteworthy multicultural team performance enhancers:

Knowledge exchange skills: over time, D&D teams develop hard and soft skills when facing and resolving problems. Balancing between the risk of overreacting to unexpected "bad news", and the risk of minimizing additional problems' impact on issues under resolution, such teams develop effective knowledge exchange skills and processes.

Diversity at its best: vivid examples of multi-disciplinarity and multiculturalism, D&D teams foster the expression of diverse perspectives in the divergence phase of the design process. Then, through a succession of divergence (generation of ideas) and convergence (elimination of ideas) phases, they use tools and techniques to eventually converge toward a shared common goal: get the product or service "out of the door".

1.3. Action researcher meets scholar in intercultural management

After the longitudinal study was completed, the action researcher confronted her experience and analysis to scholars who had not experienced the teamwork. Based on her narrative, a North American scholar studying organizational and cultural contexts conducive of innovation, suggested writing down the case and run a first analysis of multicultural dynamics as potential source of innovation. Although the cultural heterogeneity of the project team was leading to numerous discussions and variety of perspectives (as described in situation 1), the author however could not feature causal relationships between the multiculturalism of the group and the innovative nature of the project. It certainly had an influence, but no direct correlation to the innovativeness of the end product could be drawn.

The researcher then presented the case study at an international conference in international management, where she met the co-author of this chapter, a scholar from France, expert in intercultural management. She invited the author to further analyze the case with new theoretical frameworks and perspectives. In particular, she offered to analyze the case under the concept of “loosely coupled systems”, which she had identified as a theoretical perspective that could deliver a better understanding of the “diversity/consensus dilemma” multicultural teams are confronted by (Bachmann, 2006; Gibbs, 2006).

With the integration of this researcher external to the study, a second process of questions/reflection was initiated. The external researcher challenged the action researcher to rethink the process and events of teamwork, on how the events exactly occurred, on how they could be interpreted. The external researcher suggested connections to theoretical frameworks and explanations of what happened, which were in turn comforted or criticized by the action researcher.

2. ANALYSIS OF TWO INTERCULTURAL TEAMWORK SITUATIONS

We will now present, and comment, two lessons learned from this case study. Both situations illustrate how the project team has been confronted with intercultural situations that required intervention in the project plan, in a proactive and in a reactive way. Some action research tools and design-driven approaches were implemented as additional methods to resolve intercultural issues where traditional management techniques were insufficient to enhance team performance. The originality of this approach lies in borrowing tools and techniques from the discipline of design to address intercultural challenges in teams. We will first outline each event and the adopted solution, then present how reflection-in-action contributed to manage the situation before finally discussing how we theorized from this action-research process.

2.1. Situation 1: the pre-mortem exercise

Within six months, the team size grew from 8 to 25 members, and so did the team cultural heterogeneity, and the number of its asynchronous and virtual communications. Meanwhile, interdependency of tasks started to show and as a result, communication challenges increased: some requests such as sales forecasts were left unanswered or came back with partial answers; unsure of product success, some country managers in South East Asia and the UK did not want to commit on numbers. Similarly, the size of the product references was

varying depending on markets, and when leadership asked for a phased launch (starting with a limited range of references), reaching consensus on the limited range became challenging.

Adopted solution: A pre-mortem exercise involving the North American core team made of American citizen, the Asian sub-team made of citizen from Singapore, Hong-Kong and mainland China, and the United Kingdom sub-team made of citizen from the UK, was organized. Disciplines involved were engineering, ergonomics, research, marketing, and product management.

The exercise required team members to operate in culturally diverse groups of three (each group included at least one member from each sub-team described above). Six groups were formed, and the goal for each team member was to imagine *anything* that could go wrong, twelve months after the product launch. By projecting themselves into the future, in the shoes of other team members or stakeholders (clients, partners, colleagues), team members were invited to share their concerns in small groups and exert empathy over those potentially impacted by the issues. Second to this “dystopian exercise”, each small team was given time to work together, sort issues through, and present to the extended team their recommendations supported by a set of solutions to mitigate risk. If issues couldn’t be resolved by the team which expressed them, the help of the entire team was required. All the solutions were recorded. This list of issues with alternative solutions was printed on large posters that stayed during the entire length of the project on the “war room” walls. Over the course of the project, multiple references have been made to these hypothetical solutions, as foreseeable issues became reality. To sum up, all team members had closely interacted, decreased their anxiety levels and generated ideas in a non-urging context: all were better prepared to offer a wide range of solutions, when required.

Reflection-in-action: Within the design process, the team leader included a session that was not originally planned, with the aim of triggering the expression of personal or project-related challenges. As described above, the goal was to identify potential intercultural challenges *before* they happen. By grouping team members according to task and biodemographic heterogeneity, a context was offered for the expression of concerns, or fears, in the realm of small groups. In the six groups created, all recorded lists of business and personal concerns. To respect group workstyle preferences and to overcome language misinterpretations, team members were offered to reflect individually and write down their insights before regrouping; some team members chose to share and discuss issues over two or three sessions.

The process was a collaborative knowledge exchange process, within an engagement-focused diversity climate. The small teams quickly appropriated the situation and problems. Since the exercise was speculative in nature and aimed at reducing the impact of potential future conflicts, and reducing fears, it actually increased the cooperation ties between team members. The North American colleagues offered listening time and space for the colleagues in Asia and the UK to express their concerns, outside of any hierarchical, nor project timeline pressuring context. It also turned out to be a revealer of where everybody’s fears or concerns resided, without having to go through the emotional distress associated to the reality of the issue. Later on, since the teams’ lists were kept on the project room walls, it was possible to identify, when roadblocks surfaced, who was most likely going to become uncomfortable with

a given event, or occurrence. These cues allowed the team leader to seek additional support or resources, to advance the problem-solving process in a somewhat proactive way.

Theorization: The inclusion of external researcher (the second author) permitted to analyze this situation through the lens of systems theory (Orton & Weick, 1990). If the D&D team is seen as a loosely coupled system (Bachmann, 2006), then the pre-mortem exercise focused on the task to be accomplished (coupling concerning structural aspects) through social and intercultural interaction (coupling concerning cultural / social aspects). The team members were not required to agree on common values or behavior (no tight cultural coupling requirement) but rather, to adopt an empathetic perspective, which generated tighter cultural coupling.

Through the lens of design and according to the interpretation of the action researcher, the designer is often positioned as a facilitator, as emphasizing the aptitudes and activities required to ensure democratic and productive participation (Light & Akema, 2012). Recent studies focusing on the success factors of co-design projects reveal a new role for designers as “design generators”, working in the realm of collaboration to facilitate the mixing of “abstract” experts and “concrete” people to transform the way professionals work by inserting more creative design thinking (Lee, 2008). This insight leads to the following interpretation: by alternating phases of divergence in search of knowledge or ideas, and phases of convergence to select and test the most promising among them, what might represent a struggle for intercultural managers feels very familiar ground in a design context. The designer, who constantly tunes communication level between “abstract” experts and “concrete” people, develops an ability to listen to and capture every sound bite within the team, for the benefit of the various players whose inputs and insights are integrated in the formulation of potential solution sets.

2.2. Situation 2: price positioning in Asia

The Asia marketing team based in Singapore was concerned about price positioning and could not get the South East Asian markets, nor the Chinese market managers to agree on price differences between the USA and Asia. The sales organization in Asia was not ready to make any commitments on volume forecasts at this stage, which became very problematic for assessing the potential financial health of the entire project. Several meetings were planned between the American core team, the sub-teams from Asia and from the UK, the UK and American operations engineers, as well as UK and American supply managers. But no solution could be found.

Adopted solution: The Asia marketing leader based in Singapore offered to send to the Core Team a written and comprehensive document outlining their concerns, and required a discussion to follow at the American core team’s earliest convenience. The document was well articulated and documented. Some elements were to be expected, i.e. a lower price positioning, a reduced number of user options, or the desire for a saturated color palette. Other elements came in as a surprise to the Core Team: the request for a high rate of “locally manufactured components” to meet sustainability goals, as well as an aggressive launch schedule.

Reflection-in-action: The document triggered numerous concerns within the Core Team, but the Core Team leader understood that the Asian team's credibility was at stake and decided to take a step toward their requirements. The Core Team made an official, corporate announcement that the cost levels required to meet a lower price positioning in Asia could not be met, as the price gaps between various regions of the world could not exceed a certain percentage. Few days after, the information was officially shared with the extended Asian markets managers. The Asian managing director required a direct conversation with the Core Team leader, which triggered a comprehensive cost review with leaders of the other regions in Europe and the Americas. The outcome of this review became the starting point of a large sourcing assignment for the core team supply manager, which eventually yielded significant reductions in local production costs.

Theorization: Through the lens of intercultural management and the framework of intercultural team's coupling (Bachmann, 2006; Orton & Weick, 1990), the Asian sub-team increased structural coupling (in putting into writing their vision of the product for the Asian market) that caused some tighter cultural coupling (better understanding of the other's worldview through the direct conversation with the team leader). Later on, this caused still more structural coupling (through the sourcing assignment).

Culturally-ingrained business concerns, particularly when related to local markets' price positioning, are common sources of intercultural conflict that frequently lead to stalemate situations (no party wants to give in). Creating a safe place for the Asian team to talk enabled to unlock other issues unknown from the team. Moreover, inviting people to talk, and express their concerns in a written format that could be shared with executives, restored the balance in the power relations between Asia and North America. This opening happened thanks to the combination of two methods: reflection in action offered a new framework for sharing concerns, and intercultural competence created a context for safely expressing two strong requirements completely unexpected by the other culture.

3. WHAT ADVICE CAN BE GIVEN TO SCHOLARS WHO INTEND TO DO ACTION RESEARCH IN INTERCULTURAL CONTEXTS?

3.1. Can action research be led in any team or organization?

The scope and direction of a manager's reflection in action are strongly influenced, and may be severely limited, by the learning systems of the organization in which he practices (Schön, 1983). Therefore, action research is probably not appropriate in every organizational context. The organizational culture needs to be open enough to allow for successions of trial and errors, and open enough in its values and beliefs to allow the collective learning to benefit from detection and correction of errors propositions (Argyris & Schön, 1978). This case highlights action research in a D&D team, where trials and errors are part of daily activities. D&D leaders are asked to integrate recognition and correction of a perceived fault or error (Argyris & Schön, 1978) in their project activities, as primary shepherds of Toyota lean thinking model to eliminate "waste" from office processes (Womack & Jones, 1996). If the organizational culture allows for

it, and the team leader is interested in practicing action research, then there is no reason why it should not be tried.

3.2. Can any organizational actor do action research?

At the different stages of a new project, teams often need to make decisions in the absence of adequate prior information - or rules- for supporting these decisions. Some companies believe in market tests, focus groups, etc. Others will rely on their own user-centered research experiments and the intuitive drive of their creatives.

Action research does require particular prerequisite competences, and these can be learned by the organizational actor who intends to do action research: listening skills, rigor and discipline, relentless desire to find the “truth” (or the causal effects of identified symptoms), and some belief that “there always is a better way”. On a personal basis, curiosity, empathy, generosity (shared leadership), transparency (open communication is paramount), critical inquiry, and the ability to always combine advocacy when required to move on. In the context of intercultural team’s management, the two situations above outline some specific initiatives that helped overcome some challenges. The first situation highlights a proactive approach to intercultural management: offering room for team members to envision the future and discuss in small group “what might go wrong” offers a common ground for all members to express fears and concerns, in writing and in person. Small groups offered a “safe to share” context, which provided to the leader a wealth of cultural perspectives, and assessments of the project viability.

The second situation highlighted an acknowledgement from the core team that a sub-team needed the support of a larger group to pass a difficult message to their local hierarchy. Being on the constant watch of subtle signals is a competency that needs to be nurtured.

3.3. What methodology is appropriate for action research?

If they intend to develop further data analysis and publication of their research, action researchers must archive a maximum of empirical data including data “produced by organizational action” (e.g. decision reports, prototypes, accounting of other figures...) as well as process data. These include successive versions of different documents, but also notes on interaction (who, why, what? Etc.). Altogether, this is likely to result in a huge number of documents and data that will be hard to sort out and analyze later. Therefore, holding a “research diary” is highly recommended to action researchers so they can record their critical events and dates. The research diary is helpful in three aspects: 1) it provides a “short version” of the studied events and processes, 2) it helps retrieving critical data within the mass of data produced (for example sorted by the date they occurred), and 3) it triggers points of reflection on the case studied.

As described in this chapter, a daily reflection is essential. Some days it could just happen at the end of the day or the next morning, while reviewing past events or activities. Some days, it will require a larger group of contributors to reflect on more complex occurrences.

3.4. Is it helpful to collaborate with external scholars?

The collaboration with external scholars is important for action research. The action researcher experiences the empirical process from the inside, and an external researcher looks at this process with an outsiders' point of view. The combination of both can be highly beneficial. The action researcher contributes with his/her in-depth empirical knowledge and deep understanding of the situation, but which can be further revealed and developed when challenged by an external researcher.

The contribution of an external researcher is essential to provide theoretical interpretation and frameworks. As described earlier, practitioners who engage in action research often intuitively respond to situations, and the reflection-in-action benefits greatly from the influx of theories and concepts tested in other domains, or contexts. Practitioners observe and feel the same phenomena as scholars, they just do not know how to name them, nor how to connect them to existing theoretical frameworks.

Ideally, the dialogue between the action researcher and the external scholar should be led in parallel with the empirical process (Bartunek, Avenier, & Bloch Dolande, 2015). If this does not occur, discussion of the situation afterwards still can be beneficial, as it has been the case here.

CONCLUSION

This chapter has focused on the application of action research methodology to an intercultural management process. Two situations extracted from a case study presented how a practitioner has used reflection-in-action to trigger innovative answers to managerial situations, which could have led to stalemate status, or significantly delayed the project process. In both cases, the challenge was to make people from different cultures safely express their concerns, apprehensions or requests. Creating a safe context to speak up is often difficult, particularly vis-à-vis the cultural group that holds the powerful high hand in term of social categorization or language (Zhang & Guttormsen, 2015). The reflective practitioner acted as a bridge for team members from diverse cultural background and locations to access the project's daily activities, and enabled it in a transparent and highly interactive way. The chapter demonstrates that the reflection in action approach, combined with intercultural competence, has unlocked two situations which could have led to worsening intra-team communication (fear and anxiety of project failure) or to create critical conflict between the North American and Asian teams. The resolution of the "diversity/consensus dilemma" happened without compromising the diversity of thinking by triggering a safe place to speak. At the same time, consensus was achieved as a result of a fact-based debate, which integrated diverse market expectations, without compromising the product. It actually made the product better and more environmentally friendly in Asia than originally planned.

After reviewing the adopted methodology and comparing it to existing frameworks in international management, we think that action research can be led by most intercultural leaders, but not necessarily in every organization. The organizational context has to be based

on freedom of exploration and of expression, critical thinking and relentless efforts to learn and improve.

REFERENCES

- Adelman, C. (1993). Kurt Lewin and the Origins of Action Research. *Educational Action Research*, 1:1, (7-24).
- Adler N. J., Aycan Z. (2018). Cross-cultural Interaction: What We Know and What We Need To Know. *The Annual Review of Organizational Psychology and Organizational Behavior*, 5:307-33.
- Argyris, C., & Schön, D. (1978) *Organizational learning: A theory of action perspective*, Reading, Mass: Addison Wesley.
- Avenier, M. J., & Schmitt, C. (2009). Un cadre méthodologique pour des recherches tirant parti de l'expérience de praticiens de la gestion en PME. *Economies et Sociétés – série Economie de l'entreprise*, 43(2), 271–294.
- Bachmann, A. S. (2006). Melting Pot or Tossed Salad? Implications for Designing Effective Multicultural Workgroups from a Coupling Perspective. *Management International Review*, 46 (6): 721–747.
- Badrinarayanan, V. (2008). Effective virtual new product development teams: An integrated framework. *Journal of Business & Industrial Marketing*, 23: 242–248.
- Bartunek, J. M., Avenier, M. J., & Bloch Dolande, M. (2015). Dialogics and Academic-Practitioner Dialogues. In *Academy of Management Proceedings 2015* (1), 127-98. Briarcliff Manor, NY 10510: Academy of Management.
- Berg, B. (2004). *Qualitative research methods for the social sciences*, Boston. Pearson Education, 5th edition, Ch.7.195–207.
- Chua, R. Y., Roth, Y., Lemoine, J. F. (2015). The impact of culture on creativity: How cultural tightness and cultural distance affect global innovation crowdsourcing work. *Administrative Science Quarterly*, 60(2): 189-227.
- Friedman V., Antal, A.B. (2005). Negotiating Reality: a Theory of Action Approach to Intercultural Competence. *Management Learning*, 36 (1), 69–86.
- Gibbs, J. L. (2006). Decoupling and coupling in global teams: implications for human resource management. In: G. K. Stahl and I. Björkman. *Handbook of Research in International Human Resource Management*. Cheltenham, Edward Elgar Publishing: 347–363.
- Gully, S. M., Devine, D. J., & Whitney, D. J. (1995). A meta-analysis of cohesion and performance: Effects of level of analysis and task interdependence. *Small group research*, 26 (4), 497–520.
- Gustavsen, B. (2006). Theory and Practice: the Mediating Discourse. In: P. Reason and Hilary Bradbury: *Handbook of Action Research*. London: SAGE: 17-26.
- Hajro, A., Gibson, C., & Pudelko, M. (2017). Knowledge Exchange Processes in Multicultural Teams: Linking Organizational Diversity Climates to Teams' Effectiveness. *Academy of Management Journal*, 60(1), 345–372.
- Kemmis, S. (2006). Exploring the Relevance of Critical Theory for Action Research: Emancipatory Action Research in the Footsteps of Jürgen Habermas. In: P. Reason and Hilary Bradbury: *Handbook of Action Research*. London: SAGE: 94-105.
- Lee, Y. (2008). Design participation tactics: the challenges and new roles for designers in the codesign process. *International Journal of Co-Creation in Design and the Arts*, 4 (1), 31–50.

- Lewin, K. (1946). Action research and minority problems. *Journal of Soc. Issues* 2(4): 34–46
- Light A., Akama, Y. (2012). The human touch: from method to participatory practice in facilitating design with communities. In *Proceedings of the Participatory design conference: PDC2012*. Roskilde, Denmark: ACM.
- Münch, F. (2016). From Europe to North America to Asia: Overcoming the Hurdles of Interdisciplinary Multicultural Teams through a Design-Driven Process. *Management International/International Management/Gestión Internacional*, 20 (Special Issue): 38-48.
- Orton, J.D., K.E. Weick (1990). Loosely coupled systems. A reconceptualization. *Academy of Management Review*, 15(2): 203-223.
- Romani, L., Barmeyer, C., Primecz, H., & Pilhofer, K. (2018). Cross-cultural management studies: state of the field in the four research paradigms. *International Studies of Management & Organization*, 48(3), 247-263.
- Romme, A. G. L., Avenier, M. J., Denyer, D., Hodgkinson, G. P., Pandza, K., Starkey, K., & Worren, N. (2015). Towards common ground and trading zones in management research and practice. *British Journal of Management*, 26 (3), 544–559.
- Sales, A., Traver J.A., Garcia R. (2011). Action research as a school-based strategy in intercultural professional development for teachers. *Teacher and Teaching Education*, 27 (5), 911–91.
- Schön, D. (1983). The Reflective Practitioner. *How Professionals Think in Action*. New York, NY: Basic Books.
- Stahl, G.K., Maznevski, M.L., Voigt, A., Jonsen, K. (2009). Unraveling the effects of cultural diversity in teams: A meta-analysis of research on multicultural work groups. *Journal of International Business Studies*, 41(4): 690-709.
- Tadmor, C. T., Satterstrom, P., Jang, S., Polzer, J. (2012). Beyond Individual Creativity: The Superadditive Benefits of Multicultural Experience for Collective Creativity in Culturally Diverse Teams. *Journal of Cross-Cultural Psychology*, 43(3), 384-392.
- Womack J.P., & Jones, D.T. (1996). *Lean Thinking*, New York. Simon & Schuster.
- Zander L., Butler, C.L., (2010). Leadership modes: Success strategies for multicultural teams. *Scandinavian Journal of Management*, 26: 258-267.
- Zellmer-Bruhn ME, Gibson C. (2014). How does culture matter? A contextual view of intercultural interaction in groups. In M. Yuki, M. Brewer (eds.): *Culture and Group Processes*, New York: Oxford University Press: 342-402.
- Zhang W., Levenson A, & Crossley C., (2015). Move your Research from the Ivy Tower to the Board Room, *Human Resource Management*, 54 (1), 151–174
- Zhang, L.E, Guttormsen, D.S.A (2015). Multiculturality as a key methodological challenge during in-depth interviewing in international business research. *Cross Cultural & Strategic Management*, 23(2), 232–256.
- Zhu, Y. (2019). Can participatory action research empower participants in adult education studies ? *The Canadian Journal for the Study of Adult Education*, 31(1), 63-73.

AUTHORS

Fabienne Münch

Executive Director, Global Initiatives and Strategy; University of Chicago, USA
PhD Candidate in International Management
Univ. Grenoble Alpes, Grenoble INP, CERAG, F-38000 Grenoble, France

Anne Bartel-Radic

Full Professor in International Management
Univ. Grenoble Alpes, Grenoble INP, CERAG, F-38000 Grenoble, France

EIKV-Schriftenreihe zum Wissens- und Wertemanagement

Jahr	Autor/Autorin	Titel	Band
2015	Francesca Schmitt	Intellectual Property and Investment Funds	Band 1
2016	Sebastian Fontaine	The electricity market reinvention by regional renewal	Band 2
2016	Tim Karius	Intellectual Property and Intangible Assets - Alternative valuation and financing approaches for the knowledge economy in Luxembourg	Band 3
2016	Irena Hank	Emotionale Intelligenz und optimales Teaming – eine empirische Untersuchung	Band 4
2016	Pascal Berg	European Market Infrastructure Regulation (EMIR)	Band 5
2016	Dr. Sverre Klemp	Die Angemessenheit der Vergütung nach § 32 UrhG für wissenschaftliche Werke im STM-Bereich	Band 6
2016	Lars Heyne	Immaterialgüterrechte und Objektreplication: Juristische Risiken und Lösungsmöglichkeiten bei der Vermarktung von 3D-Druckvorlagen	Band 7
2016	Torsten Hotop	Äquivalenzinteresse im Erfinderrecht	Band 8
2016	Christian Wolf	Zur Eintragungsfähigkeit von Geruchs- und Hörmarken	Band 9
2016	Nadine Jneidi	Risikofaktor Pflichtteil - Grundlagen und Grenzen der Regelungs- und Gestaltungsmöglichkeiten von Pflichtteilsansprüchen bei der Nachfolge in Personengesellschaften	Band 10
2016	Meika Schuster	Ursachen und Folgen von Ausbildungsabbrüchen	Band 11
2016	Julie Wing Yan Chow	Activity Based Costing - A case study of Raiffeisen Bank of Luxembourg	Band 12
2016	Peter Koster	Luxembourg as an aspiring platform for the aircraft engine industry	Band 13
2016	Stefanie Roth	The Middle Management – new awareness needed in the current information society?	Band 14
2016	Alexander Fey	Warum Immaterielle Wirtschaftsgüter und Intellectual Property die Quantenteilchen der Ökonomie sind	Band 15
2016	Daniel Nepgen	Machbarkeitsstudie eines Audioportals für Qualitätsjournalismus. Eine empirische Untersuchung in Luxemburg	Band 16
2016	Niklas Jung	Abolition of the Safe Harbor Agreement – Legal situation and alternatives	Band 17
2017	Marco Pate	Kriterien zur Kreditbesicherung mit Immaterialgüterrechten anhand der Finanzierungsbesicherung mit Immobilien	Band 18

Jahr	Autor/Autorin	Titel	Band
2017	Patrick Matthias Sprenker	RAIF – Reserved Alternative Investment Fund – The impact on the Luxembourg Fund Market and the Alternative Investment Fund landscape	Band 19
2017	Sebastian Fontaine	Quo vadis Digitalisierung? Von Industrie 4.0 zur Circular-Economy	Band 20
2017	Andrea Dietz	Anti-Money Laundering and Counter- Terrorist Financing in the Luxembourg Investment Fund Market	Band 21
2017	Christophe Santini	Burn-Out / Bore-Out Équivalences, similitudes et différences impactant la vie socio-économique des personnes concernées	Band 22
2017	Johanna Brachmann	Ist das Arbeitnehmererfindungsrecht erneut reformbedürftig? - Ein Rechtsvergleich zwischen Deutschland und Österreich, Schweiz, USA, Großbritannien	Band 23
2017	Nadine Allar	Identification and Measurement of Intangibles in a Knowledge Economy - The special relevance of human capital	Band 24
2018	Alexander Vollmer	Überwachung von ausgelagerten Funktionen und Kompetenzen in der luxemburgischen Fondsindustrie	Band 25
2018	Claudia Lamberti	Women in management and the issue of gender- based barriers - An empirical study of the business sector in Europe	Band 26
2018	Désirée Kaupp	Corporate culture - an underestimated intangible asset for the information society	Band 27
2018	Romain Gennen	Die automobile (R)Evolution – das automobile Smartphone	Band 28
2018	Sven Kirchens	TVA - Introduction du mécanisme de l'autoliqui- dation dans le secteur de la construction au Luxembourg ? Analyse et Propositions	Band 29
2018	Lisa Schreiner	The Effects of Remuneration and Reward Systems on Employee Motivation in Luxembourg	Band 30
2018	Alina Bongartz	Der Einfluss der Kundenzufriedenheit auf den Unternehmenserfolg - die Wirkung von Value Added Services	Band 31
2019	André Reuter Thomas Gergen (Hg)	Studien zum Wissens- und Wertemanagement Investment, Gesundheitswesen, Non-Profit- Organisationen, Datenschutz und Patentboxen	Band 32