



Adaptation internationale des échelles de mesure entre universalisme et culturalisme : application à la mesure de l'environnement de l'entreprise *

Dominique Besson

Maître de conférences

IAE de Lille

104 Av. du peuple belge, 59000 Lille

Tél : (+33) (0)6.09.73.10.32

E-mail : Dominique.Besson@iae.univ-lille1.fr

dominique.besson13@wanadoo.fr

Slimane Haddadj

Maître de conférences

IAE de Lille

104 Av. du peuple belge, 59000 Lille

Email : Slimane.haddadj@wanadoo.fr

Résumé

Dans le cadre de recherches en stratégie impliquant une dimension internationale, la question de l'utilisation d'échelles de mesure ayant une validité dans plusieurs pays se pose. On se rend compte que souvent des chercheurs utilisent des instruments de mesure venus des Etats Unis sans se soucier de leur fiabilité et validité. Plus précisément, faute d'une réflexion sur les construits et de test sur les énoncés des items, les écarts de validité ou les différences de facteurs dans les résultats de l'administration de mêmes échelles de mesures dans des pays différents sont souvent expliqués par des différences culturelles irréductibles. Une démarche méthodologique plus rigoureuse doit dépasser le constat de la différence culturelle, souvent prétextée pour éviter de rechercher la fiabilité et la validité d'un instrument de mesure. Une telle démarche permet d'utiliser les apports des tenants de chacune des deux thèses du débat entre universalisme et culturalisme dans l'adaptation des échelles de mesures.

Après une réflexion sur ces questions, ce travail adapte en français une échelle de mesure de l'environnement de la firme à partir de travaux nord-américains (Miller, Friesen, 1982, 1983), en appliquant une méthode rigoureuse. L'apport de cette communication est triple. Cette échelle de mesure pourrait devenir un outil de mesure pour les recherches qui se font en

* Les auteurs remercient Jean-Pierre Neveu, professeur à l'Université de Limoge, Dominique Crié et Richard Ladwein, maîtres de conférences à l'Université de Lille et Charles-Henri Beyssere des Horts, Professeur HEC, pour leurs commentaires sur les premières versions de cet article. Les éventuelles erreurs et insuffisances restent, bien entendu, de notre seule responsabilité.



France. Traduite dans plusieurs langues, cette recherche pourrait permettre dans le cadre de recherches comparatives au niveau international de disposer de données construites sur des instruments fiables et validés, et ainsi permettre aux chercheurs de pouvoir appréhender les mêmes construits sur le plan conceptuel dans plusieurs pays. Enfin, le travail que nous proposons, inspiré du paradigme de Churchill (1979) sur le plan de sa méthodologie pourrait servir de guide aux chercheurs en gestion qui voudraient développer des échelles de mesure.

Mots-clé

Méthodologie, GRH, international, environnement, stratégie, facteurs, construits, culture, analyse comparative.

Abstract

In the international Business Administration research, the question of using scales of measure that have validity in several countries should be asked. In fact very often, researchers use scales of measure coming from the U.S., but they did not take into account the reliability and the validity of the measure in their translated scale. In this case, it is important to overcome the idea of the cultural difference, which is invoked in order to avoid to look for reliability and validity.

Based on problems related on reliability and validity, this article proposes to develop a French scale measure from scales that already have been used in North America. The interest of this article is to provide a French questionnaire for researches in France. Moreover, translated in several languages, the questionnaire could be used in international researches that are looking to measures constructs that possess reliability and validity. At last, inspired from the paradigm of Churchill (1979), this article provides the opportunity to use a rigorous methodology in order to develop scales of measure.

Keywords

Methodology, HRM, International, Environment, strategy, Factors, Constructs, Culture, Comparative Analysis.



Adaptation internationale des échelles de mesure entre universalisme et culturalisme : application à la mesure de l'environnement de l'entreprise

1. INTRODUCTION

Il existe deux principales optiques pour aborder la contextualité internationale en gestion. La première manière relève plutôt de l'analyse des jeux d'acteurs et est conçue pour répondre à des contextualités locales en cherchant à découvrir les régulations qu'élaborent les acteurs internes de l'organisation (« environnement interne »). Les questions culturelles y sont souvent convoquées (Schein, 1990, Hofstede, 1987). La deuxième est celle qui lie les domaines des sciences de gestion aux contingences qui se situent à l'extérieur de l'entreprise, contingences qui contribuent à articuler l'organisation à son environnement. Dans cette optique les références relèvent plutôt des théories de la contingence dans la détermination de la stratégie. L'entreprise chercherait à améliorer le rapport entre l'environnement et la stratégie de l'organisation, ce qui serait source d'avantages concurrentiels. Si le processus d'internationalisation des entreprises va croissant dans une économie qui se globalise, aucune des fonctions de gestion n'échappe à ce processus, et elles doivent répondre à des exigences contradictoires de globalisation et de contingence locale liée au pays dans lequel l'entreprise décide de s'implanter (Lecomtes et Forgues, 2000).

Nous nous situons dans le cadre de cette hypothèse selon laquelle l'environnement de l'entreprise peut être un élément de contingence qui structure les stratégies des entreprises afin d'améliorer leurs performances organisationnelles. L'objectif principal de cette recherche est d'évaluer les aspects psychométriques interculturels d'une l'échelle de mesure de l'environnement de l'entreprise. Le choix d'une échelle de mesure validé dans un pays donné ne suffit pas à en garantir la validité dans d'autres pays. La validité des instruments de mesure peut affecter la pertinence et la qualité de toute décision organisationnelle, notamment pour les entreprises qui se diversifient géographiquement, et qui désirent recueillir des informations fiables dans les pays où elles chercheraient à développer leurs activités. Par exemple, Davis, Douglas et Silk (1981), dans un article déjà ancien mais toujours d'actualité, ont souligné que les différences entre deux moyennes dans la distribution d'une variable peuvent ne pas être significatives si les écarts de fiabilité concernant ces moyennes sont assez importants. La question est aussi importante pour les chercheurs amenées à faire des recherches sur le



comportement des entreprises, notamment lors de comparaisons interculturelles, où l'échelle de mesure serait utilisée dans des pays qui ne l'auraient pas encore testée. Pour ces deux raisons, il est donc important d'établir les propriétés psychométriques de ces instruments de mesure à travers plusieurs pays (Netemer et al. 1991).

Répondant à ces deux critères, l'échelle de mesure de l'environnement de l'entreprise issue des travaux nord américains de Friesen et Miller (1982, 1983) a été choisie. Cette échelle de mesure, qui n'a jamais été traduite et validée en France, pourrait devenir un outil de mesure pour les recherches comparées entre la France et l'Amérique du nord, ou pour les recherches en France qui veulent utiliser les mêmes construits que ceux qu'ont développés Friesen et Miller (1982, 1983). Pour ce faire, cette échelle de mesure a été validée en France et aux Etats-Unis à partir du paradigme de Churchill (1979) et des méthodes dites « de seconde génération » (Valette-Florence, 1988) qui s'appuient sur les travaux de la psychométrie concernant en particulier la distinction entre fidélité et validité. Le travail que nous proposons, sur le plan de sa méthodologie, pourrait servir de guide général aux chercheurs en gestion qui voudront développer des échelles de mesure ou traduire des échelles de mesure à partir de travaux de recherche réalisés dans une langue différente de la leur, soit dans le cadre de comparaison internationale, où d'étude en France à partir de l'importation de concepts (quand cela a un sens), construits, mesurés et validés de manière rigoureuse.

2. LE « CULTUREL » ET « L'UNIVERSEL » DANS L'ADAPTATION INTERNATIONALE DES ECHELLES DE MESURE

Pour comprendre de manière appropriée les contextes à la fois nationaux et internationaux, la question de l'utilisation d'échelles de mesure ayant une validité dans plusieurs pays se pose. De manière critique, on pourrait poser la question de la manière suivante : comment rendre compte d'une évaluation fiable et pertinente des différences et similarités culturelles, lorsque dans chacun des pays où le questionnaire est administré, les construits ne sont jamais identiques, et donc que la fiabilité et la validité des construits ne sont pas données *a priori* ? Quelles que soient les remarques critiques que l'on peut apporter aux travaux de G. Hofstede, notamment sur le plan méthodologique (Tayeb, 2001), de nombreuses études, y compris de type qualitatif, sont convergentes avec ses résultats : les pratiques (au sens de Hofstede, 1993, p.24) et les perceptions sont fortement variables selon les contextes nationaux (Hofstede, Triandis, et Smith, 2001, Janssens et al., 1995).



Pour mesurer les perceptions des individus dans des contextes nationaux différents, les chercheurs transposent de manière fréquente des instruments de mesure venus des Etats-Unis, sans suffisamment se soucier de leur fiabilité et de leur validité dans le nouveau contexte d'application. Les chercheurs partent de l'idée que les instruments de mesure sont universels, et qu'ils fonctionneront de manière efficace à la fois dans le pays où ces instruments de mesure ont été conçus et dans les pays où ils seront utilisés. Peut-on considérer que les échelles nord-américaines représentent des échelles globales supra-nationales ?

Sans se soucier de la différence culturelle lors de la *construction* des échelles de mesure et de l'*administration* des questionnaires, des chercheurs font le constat de la différence culturelle *au moment où ils analysent les résultats de leurs données*, et apportent une solution pratique en réinterprétant les dimensions trouvées à partir de leur données. Cette réinterprétation des dimensions implique qu'elles ne reflètent plus les construits initialement convoqués. Ceci peut être compris comme une conséquence des différences culturelles. Mais cette démarche est contestable en ce qu'elle relève dans certains cas d'une tautologie. Si dans le processus de traduction des indicateurs de mesure le contexte culturel n'est pas pris en compte, c'est la manière même dont ont été élaborés ces indicateurs qui peut être remise en question, faute d'un processus méthodologique rigoureux. Les différences dans les dimensions des construits ne relèveraient alors pas de différences culturelles, mais d'une mauvaise (re)construction des indicateurs. Il ne s'agit pas que de simples problèmes de traduction, mais bien de construction des items en prenant en considération(s) leurs liens nomologiques (ceux qui relèvent du sens que les répondants leur attribuent les uns par rapport aux autres).

Il existe un débat profond sur cette question. Chacun observe que les contraintes juridiques, culturelles et sociales peuvent grandement différer. En revanche, quelle est la portée de ces différences sur les concepts qui structurent les décisions et les pratiques ? Le courant « culturaliste » en gestion estime que les différences ont des effets significatifs, particulièrement dans tout ce qui « oriente l'action des hommes » (Iribarne, 1987, p.9). A l'opposé, d'autres estiment que les concepts, les ressorts profonds des actions au-delà des différences culturelles au niveau des pratiques, restent souvent les mêmes y compris dans les domaines du management des Ressources humaines et des problèmes d'attitudes et de comportements (Igalens, Roussel, 1998). Nous relevons ci-dessous quelques arguments du débat à partir d'exemples de recherches appliquées.

Dépasser ce débat implique d'entreprendre un travail méthodologique pour chaque utilisation d'échelle de mesure d'origine étrangère, afin de pouvoir mesurer avec précision les réponses



des individus à des construits précis. Faute de ce travail, il est possible d'aboutir à un mauvais instrument de mesure, c'est-à-dire qui ne présente pas les niveaux minimums de fiabilité et de validité, procure de mauvaises données et, par conséquent, conduisent à des résultats erronés.

2.1. LES DIFFERENCES CULTURELLES A L'ORIGINE DES ECARTS DE MESURE

Certains chercheurs affirment que les échelles de mesure sont fortement connotées culturellement et qu'elles nécessitent des adaptations contextuelles liées à la culture qui dépassent le simple problème de la traduction (Whiting, 1968, Douglas et Craig, 1983). Dans la mesure où les propriétés psychométriques varient d'un pays à l'autre, on peut en conclure que l'échelle n'est pas fiable et pas valide et qu'elle nécessite de prendre en considération des aspects culturels.

Earley et Singh (1995) ont proposé une typologie des approches en recherche interculturelle, et distinguent quatre principaux modèles (unitaire, gestaltiste, réduite et hybride). La première et dans une certaine mesure la seconde relèvent du « culturalisme ».

La forme unitaire (*Unitary Form*, p.330) est ce que d'autres auteurs appellent l'approche Emic (à partir de la description sémiotique) : l'accent est mis sur une approche « unitaire » d'une seule culture, dans ses propres termes et en utilisant ses propres construits. Il s'agit d'utiliser ceux-ci sur l'occurrence d'un phénomène précis, dans un contexte culturel donné. Dans la forme gestaltiste (*Gestalt Form*, p.330), il s'agit d'analyser un système culturel comme un tout, au lieu de le décomposer en parties. Mais cette approche est cependant interculturelle : les liens entre les construits sont, par hypothèse, dérivés de principes généraux et non de chacune des cultures analysées.

Des auteurs ont ainsi souligné l'ethnocentrisme dans les comportements économiques de consommation (Shimp, Sharma, 1987, Netemeyer et al., 1991). L'analyse de Parameswaran et Yaprak (1987) a montré que les fiabilités des instruments de mesures dépendaient de manière statistiquement significative de certaines variables liées aux différences culturelles (les attitudes, la nationalité des répondants aux questionnaires, et le pays d'origine des produits et services (ce que l'on peut étendre aux méthodes de gestion). Leur méthodologie utilise de manière très poussée les recommandations de double traduction inversée de Whiting (1968). Ce type de travaux souligne la nécessité de porter une grande attention à la validité nomologique des questionnaires (sens des items selon les différentes catégories de répondants). Une tentative de validation dans des pays de l'Europe de l'Est de l'échelle de l'ethnocentrisme des consommateurs CETSCALE de Shimp, Sharma, (1987) a montré que



cette échelle ne pouvait pas leur être appliquée, du fait de différences culturelles dans les comportements (Lindquist et al., 2001).

Dans un tout autre domaine d'application, celui de la psychologie sociale, Smith et al. (1991) ont testé la traduction d'un questionnaire anglais sur un échantillon japonais, appliqué aux rythmes circadiens. Les résultats de l'adaptation sont globalement mauvais, du fait des insuffisances des traductions. Les auteurs ont testé le questionnaire à partir d'une traduction dans un seul sens. Les faibles qualités psychométriques du questionnaire traduit (établies par comparaison des poids factoriels et des tests de Chi²) les amènent alors à opérer une traduction inversée, et à constater les nombreux problèmes de non équivalences entre le sens des items en anglais et en japonais. Ils en concluent qu'il est impossible de séparer les problèmes proprement linguistiques de traduction des aspects culturels sémantiques. Mais la rigueur des outils statistiques utilisés pour tester la comparabilité de l'échelle de mesure n'a été utilisée qu'après la construction du questionnaire, et n'a pas été intégrée dans la construction même de celui-ci (Cf. plus bas les recommandations de Venkatraman et Grant, 1986, sur les liens entre constructions des relations théoriques des construits et celles de leurs opérationnalisations respectives).

Certains travaux sont conçus, dès l'origine, pour valider des hypothèses sur les différences culturelles irréductibles à l'échelle de mesure. Janssens et al. (1995) utilisent un questionnaire pour tester le mode de perception par les salariés dans différents établissements internationaux d'une même multinationale d'une politique de sécurité formalisée au niveau global de la firme. Leurs traitements statistiques sont conçus pour valider leur hypothèse centrale : les différences dans les dimensions culturelles nationales au sens de Hofstede expliquent les écarts statistiques constatés. Triandis (2001) estime même que l'utilisation de procédures quantitatives ne seraient possibles que si la distance culturelle entre la culture du chercheur et celle du terrain interrogé ne serait pas trop forte (p.15). De plus, les cultures nationales ne pourraient être décomposées en dimensions permettant de tenir compte, dans des approches quantitatives, des facteurs de culture nationale expliquant des différences dans les construits (Tayeb, 2001). Il serait d'ailleurs aussi impossible de dépasser les biais culturels des chercheurs eux-mêmes (Ibid).

Si tel était le cas, il faudrait recréer un instrument de mesure purement adapté au contexte dans lequel l'étude doit être réalisée. C'est une procédure longue et difficile car elle nécessite de bien comprendre les contextes dans lesquels on cherche à administrer le questionnaire pour trouver à travers les construits convoqués des items qui soient équivalents afin d'établir les



propriétés psychométriques des instruments de mesures qui seront utilisés dans plusieurs pays. Les chercheurs proches de cette optique montrent une grande attention aux méthodes d'adaptation des questionnaires (Whiting, 1968). Ces travaux relèvent du principe « toutes choses égales par ailleurs ». Mais ils rendent difficiles les comparaisons internationales.

2.2. L'APPROCHE « UNIVERSALISTE »

D'autres chercheurs s'opposent à cette vision culturaliste, et affirment que les échelles de mesure ont un caractère universel (Nyeck, Paradis, Xuereb, Chenat, 1996). A cet effet, il est fréquent que des échelles de mesure utilisées dans un contexte nord-américain soient simplement traduites pour être reprises dans un contexte différent (Ibid). Toujours d'après ces auteurs, cette attitude des chercheurs dénote un certain ethnocentrisme. Pour les chercheurs qui adoptent cette vision universaliste, la connotation culturelle tend à disparaître du fait de la globalisation croissante des marchés. Par exemple Levitt (1983) affirment que les goûts, les besoins, et les styles de comportement des consommateurs tendent à s'homogénéiser.

En terme d'approches cross-culturelle, il s'agit ici de la *forme réduite* (Reduced Form) dans la typologie de Earley, Singh (1995, p.331). Celle-ci décompose le système en composantes pour mieux comprendre le fonctionnement des processus dans chacune de ces composantes. Les relations entre construits sont étudiées sans tenir compte d'éventuelles autres relations avec d'autres construits. De plus, les hypothèses sur ces relations sont fréquemment dérivées d'une culture et testées dans une autre. Enfin, et par conséquent, les relations entre construits ne sont pas analysées en termes du système culturel global dans lesquels ils apparaissent, mais à partir des aspects spécifiques propres aux construits analysés.

Dans une étude des méthodes de gestion et des comportements dans des établissements américains et japonais de finance internationale, Zaheer (1995) aboutit à un constat nuancé, mais clairement « anti-culturaliste ». D'après son étude, la culture s'efface devant les pratiques spécifiques à l'organisation, quel que soit le lieu d'implantation. Ainsi, les entreprises auraient intérêt à ne pas systématiquement adopter les pratiques locales. Plutôt que de chercher à ressembler aux firmes locales, à imiter leurs pratiques et à s'accommoder à la culture du lieu d'implantation, elles auront de meilleurs résultats en renforçant, localement, leur spécificité provenant de leur firme étrangère d'origine.

Les résultats du travail « historique » de Davis et al (1981) sont plus nuancés. Leurs tests de fiabilité et de validités ne confirment pas l'hypothèse de différences culturelles ayant des effets statistiques significatifs sur les écarts entre échantillon culturellement différenciés. Leur



première conclusion semble claire : Sur l'ensemble des variables dégagées, « *No systematic tendency for any particular national sample or language group to exhibit high or low reliability consistently across several types of measurements or across several different variables of the sale type* » (p.108).

Mais leurs travaux ne valident pas de manière univoque la thèse de l'universalité des concepts. Un raffinement statistique dans le traitement des variables en distinguant les types de construits les amène à une deuxième conclusion moins affirmative, qui complexifie leurs résultats d'une manière intéressante : « *The incidence of reliability differentials appeared to vary according to the type of variable, being less likely to occur for hard variables like demographic characteristics and more likely to occur for soft variables such as task/decision involvement and life-style/psychographic factors.* » (Ibidem, p.108)

Le débat débouche ainsi sur la discussion des outils selon le type de construit abordé. Il ne s'agit pas de choisir une option *a priori*, mais de discuter des méthodes permettant de construire de manière rigoureuse les échelles de mesure.

2.3. LE DEBAT SUR LES METHODES D'ADAPTATION DES ECHELLES DE MESURE

Ce débat est très présent dans les réflexions méthodologiques des chercheurs en marketing international, particulièrement dans les difficultés d'établir des équivalences cross-culturelles (Usunier, 1992, Urien, 2000). Dans une revue de la littérature en marketing international sur ce point, Usunier (1990) dégage cinq propositions méthodologiques dont nous avons tenu compte dans notre travail d'adaptation et de validation de l'échelle de mesure de l'environnement stratégique. Les cinq propositions sont les suivantes :

Proposition 1 : « Si le type de données recherchées et les procédures de recherche sont considérées comme d'application générale (i.e. indépendante du contexte culturel, les principales différences entre la recherche marketing domestique et cross-culturelle se trouvent dans les difficultés pour établir l'équivalence aux différentes étapes du processus de recherche. » (95) ; Proposition 2 : « Des instruments de recherche marketing adaptés à chaque culture nationale (Approche « Emic ») présentent une meilleure validité du point de vue des construits qu'ils représentent, que des tests applicables directement à plusieurs cultures (Approche « Etic » ou « Culture-free test »). De ce fait, leur comparaison est plus difficile mais aussi plus fiable. » (98) ; Proposition 3 : « les techniques de traduction, même les plus achevées techniquement, peuvent s'avérer incapable d'atteindre la complète comparabilité linguistique des données. » (102) ; Proposition 4 : « Des divergences dans les styles de



réponses à travers les pays peuvent introduire des biais dans les informations collectées et fixer des limites étroites aux possibilités de comparaison directe. » (103) ; Proposition 5 : « la représentativité et la comparabilité des échantillons de recherches marketing cross-culturelles peuvent être atteints dans de meilleures conditions en utilisant des échantillons (et techniques d'échantillonnage) différents, qui produisent des niveaux équivalents de fidélité, plutôt qu'en utilisant la même procédure avec tous les échantillons. » (107).

La première proposition d'Usunier est particulièrement pertinente pour l'objet de cette étude. Dans sa revue de littérature, Usunier montre la complexité dans l'établissement des équivalences.

Les domaines d'équivalences peuvent être très nombreux (Douglas, Craig, 1984), et il devient extrêmement difficile de s'assurer de bonnes qualités psychométriques d'enquêtes qui évitent les biais qui peuvent découler de chacun des domaines d'équivalence (Douglas et Craig en déterminent 14, chacun d'entre eux pouvant être décomposé en plusieurs niveaux d'équivalences). Usunier (1990) fait référence à des auteurs qui ont souligné « qu'il est plus facile d'obtenir une équivalence de mesure entre des variables socio-démographiques qu'entre des variables psychographiques » (97). Ceci est en convergence avec l'étude de Davis et al (1981) dans un domaine pourtant étranger au marketing. Usunier souligne ensuite que « la nature du construit, aussi bien que la méthode pour l'opérationnaliser peuvent être deux causes d'infidélité comparative de la mesure. A la limite, deux mesures fidèles dans leur contexte national respectif, peuvent être infidèles en terme de comparaison entre les deux pays, dans la mesure où les deux construits qu'elles sont censées représenter ne sont pas directement comparables. » (Usunier, 1990, 97).

Grunert et al. (1993), précisent la spécificité de la question des équivalences. Les différences culturelles interviennent-elles au niveau du réel, c'est-à-dire sur le construit à mesurer lui-même, ou bien sur sa « mise en mot », c'est-à-dire au niveau de l'expression langagière du construit ? les précautions méthodologiques que nous utiliserons répondent à cette distinction. Leur adoption permet d'établir des adaptations d'échelle de mesure qui éliminent au maximum les effets langagiers, et permettent alors de postuler l'identité des construits. Il convient de garder à l'esprit que toute comparabilité reste imparfaite (Usunier, 1992 et Marchetti et Usunier, 1990).

La deuxième proposition d'Usunier nous a aussi été très utile, qui concerne le problème des équivalence conceptuelle et équivalence métrique et de l'hétérogénéité du sens. Dans la



discussion de sa proposition, Usunier n'est pas aussi catégorique que l'énoncé de la proposition elle-même. Là aussi, l'important est la rigueur méthodologique dans la construction des construits. « La validité d'une échelle psychométrique dans une étude cross-culturelle est affectée par les problèmes d'équivalence métrique et par l'hétérogénéité possible du sens. Pras et Angelmar (1978) ont fait une comparaison d'échelle à base de différentiels sémantiques, en français et en anglais. Ils montrent que des difficultés peuvent survenir lorsque l'on essaie de déterminer les équivalents lexicaux des descripteurs verbaux (adjectifs en l'occurrence) pour l'échelle commune. (...) La méthode suggérée par Pras et Angelmar conduit au rejet de l'équivalence de définition *a priori* sur la base du sens des mots utilisés dans les instruments de mesure en langage source. Ils préconisent une méthode décentrée (Campbell et Werner, 1970), ce qui signifie en clair de construire des échelles valides et fidèles pour tous les pays où la recherche est effectuée. » (Usunier, 100-101). Ce principe nous a mené à ne pas nous contenter d'une traduction « techniquement achevée », mais à donner une importance prioritaire à la fiabilité et à la validité de l'échelle de mesure en anglais et en français (y compris en comparant ces critères pour les deux versions administrées du questionnaire, cf. ci-dessous).

Ceci renforce d'ailleurs la proposition de Douglas et Craig (1984) d'une démarche hybride : comme les équivalences dans les significations posent des limites indépassables, il convient d'appliquer des méthodes rigoureuses pour ne pas faire de contresens. C'est pourquoi, sans entrer dans une position tranchée, à partir du résultat de Davis et al. (1981), des réflexions de Tayeb (2001) et Parameswaran et Yaprak (1987), il est possible d'imaginer une position intermédiaire. C'est d'ailleurs le cas de la plupart des auteurs qui pensent que les différences culturelles sont significatives et doivent être prises en compte dans la construction des outils de mesures (Green et White, 1976, Forgues, 1995).

Dans le domaine du Management, l'approche hybride (Hybrid Form) tendrait à devenir dominante (Earley, Singh, 1995, p.332). Elle cherche à mixer des caractères de la forme gestaltiste avec certains de la forme réduite. Les objets d'analyse sont déterminés à partir de l'analyse gestaltiste des systèmes afin de déterminer les aspects importants à analyser. Mais ces aspects ne sont pas conçus comme spécifiques à un système : sont analysés en priorité ceux qui semblent exister à travers les différents systèmes. Il est considéré que les construits et les relations entre eux peuvent être analysés de manière séparée d'approches globalisantes, mais l'intégration des relations dégagées par la recherche dans les systèmes existants n'est *a priori* ni linéaire ni additive. Enfin, si les relations spécifiques entre construits sont



interprétées comme des sous-systèmes réduits, la référence au système global n'est jamais écartée.

Earley et Singh recommandent la forme hybride, pour plusieurs raisons (p.334 sq.) : pragmatique, elle permet de construire et d'explorer des relations complexes qui seraient fortement interdépendantes et dépendantes du contexte.

L'approche hybride n'est pas limitée aux travaux qualitatifs, et les recommandations de recherche que l'on peut en dégager s'appliquent aussi aux études quantitatives. Elles soulignent l'importance à accorder aux soins pour construire les échelles de mesure dans ce cas. Un bon exemple en est l'étude sur les tensions de rôle au travail (conflits, ambiguïté et « surcharge » de rôles *-role overload*) dans 21 pays, et présentée par vingt-et-un auteurs au *Research Forum on International and Intercultural Management Research*, reprise dans le numéro spécial qui y est consacré dans l'*Academy of Management Journal* et que citent Earley et Singh comme innovante et représentative de l'approche hybride.

L'objectif de Peterson, Smith, et al. (1995) consiste à mesurer les liens entre les facteurs dégagés de la mesure de trois caractères de la tension de rôle (*role stress*) en organisation, que sont le conflit, l'ambiguïté, et la surcharge (*overload*) de rôle, avec les quatre dimensions de la culture nationale au sens de Hofstede (« première version) (distance hiérarchique, individualisme/ collectivisme, masculinité/ féminité, contrôle de l'incertitude).

La méthodologie utilisée est directement en rapport avec la notre. Les auteurs partent de deux questionnaires existants en anglais (l'un américain, l'autre indien) qu'ils adaptent en deux temps. Les données ont été collectées sur un échantillon homogène de cadres intermédiaires (*middle managers*). Dans chaque pays, les chercheurs locaux, expérimentés, ont effectué des traductions rétroactives et / ou parallèles.

Dans un premier temps, les auteurs montrent dans plusieurs pays de faible « fit » suite à une analyse confirmatoire sous LISREL (les items sont fixés pour déterminer chacun des trois axes), voire pas de résultat possible du tout.

Ils mènent alors une épuration de l'échelle de mesure en éliminant les items qui avaient des pondérations différentes selon les différents pays, en fonction des structures factorielles (ne restent que 13 items sur les 31 de départ). En deuxième temps, ils peuvent opérer une analyse confirmatoire multi-échantillon sur un regroupement des différents pays en trois ensembles. Les résultats dégagés sur l'échelle de mesure de la tension de rôle ont ainsi pu être fiables et valides.



Cette étude relève clairement d'une approche hybride. Le résultat global de l'étude des corrélations est plutôt « emic » : la tension de rôle varie significativement plus selon les différents pays que selon les différences de facteurs organisationnels et démographiques. Mais leur traitement méthodologique de l'échelle de mesure est résolument de l'ordre d'une approche « etic » mesurée : ils ont construit leurs échelles de mesure de manière à disposer d'un outil utilisable à travers les différents pays (ou plutôt cultures, car ils ont opéré un regroupement des pays en trois zones supposées culturellement relativement homogènes au sens de Hofstede).

3. AU-DELA DU DEBAT : LES CONDITIONS DE VALIDITE D'UNE ECHELLE DE MESURE

Il n'est pas forcément nécessaire de prendre parti dans ce débat pour mener une recherche utilisant une échelle de mesure. Que ce soit dans le cadre d'une recherche dans un contexte local ou bien dans le cadre d'un contexte international, une échelle de mesure doit posséder fiabilité et validité pour ne pas affecter la qualité des construits que l'on mesure. Comme le soulignent Evrard et al. (1993), lors d'une recherche, il est possible de rencontrer quatre situations : les résultats sont précis et vrais, les résultats sont précisément faux, les résultats sont grossièrement exacts, les résultats sont grossièrement faux. Le cheminement que propose Churchill (1979) est de se rapprocher au maximum de la vraie valeur en s'assurant de la fiabilité et de la validité des instruments de mesure que l'on convoque.

Les recherches réalisées dans un cadre international n'échappent pas à cette règle et doivent suivre le même cheminement. Il existe une littérature assez vaste dans le domaine des recherches internationales qui indique l'importance d'établir la fiabilité et la validité interculturelles des instruments de mesure avant de les utiliser pour mesurer des phénomènes à l'étranger ; si elles ne sont pas établies, les résultats obtenus seront illusoire.

Pour éviter d'être dans cette situation, un processus méthodologique rigoureux doit être suivi par le chercheur pour arriver à la fiabilité et à la validité des résultats. D'une manière pessimiste, Evrard et al (1993), en citant une recherche qui a été réalisée sur 70 programmes en sciences sociales, concluaient que la variance des résultats obtenus peut s'expliquer pour 42% par le phénomène que l'on cherche effectivement à mesurer, pour 26% par l'erreur systématique, et pour 32% par l'erreur aléatoire. Lorsque l'on utilise des questionnaires pour lesquels aucune fiabilité ou validité n'est prise en considération dans le pays où l'on veut l'utiliser, sous prétexte qu'elle a été calculée dans le pays d'origine, on risque de réduire



d'autant la variance des résultats liée au phénomène que l'on cherche à mesurer. Dans le cas spécifique du management stratégique, Venkatraman et Grant (1986) font un constat encore plus critique. Ils relèvent les nombreuses insuffisances dans les démarches d'opérationnalisation des construits théoriques des chercheurs en stratégie. Ceux-ci, d'abord, développent trop vite des prescriptions. Venkatraman et Grant, citant ici Mintzberg, soulignent le manque de prudence préalable à la prescription, qui consiste à approfondir l'investigation du réel (p.72). Les recherches utilisent très fréquemment des échelles nominales et simple-item, qui comportent de nombreuses limites. Facteur aggravant, la littérature montre, en général, une évaluation inadéquate de la validité de construit (pp.78-82).

Certaines précautions méthodologiques ont été élaborées dont la synthèse la plus utilisée est celle de Churchill (1979). Elles devraient théoriquement être appliquées par l'ensemble des chercheurs qui utilisent les questionnaires dans le cadre de recherches nationales ou internationales. Comme le disent Venkatraman et Grant (1986), dans leur étude critique de la littérature de recherche opérationnelle en stratégie, « plus d'attention aux qualités des mesures va certainement diminuer la probabilité que des vraie relations [entre construits] seront rejetées du fait de mesures déficientes » (p.72, notre traduction).

Ces auteurs conseillent d'être très rigoureux dans l'analyse de la validité de construit (pp.83 sq.). Citant Nunnally, ils relèvent que les sciences avancent à partir de construit et rarement à partir de variables spécifiques observables et que les relations entre construits dépendent de manière critique de la correspondance entre construits théoriques et leurs opérationnalisations respectives. Ils soulignent aussi que la rigueur dans l'analyse des mesures permet de diminuer les ambiguïtés sémantiques, courantes en stratégie (mais aussi, nous l'avons vu, dans d'autres domaines). C'est pourquoi ils estiment que la mesure des construits est une étape essentielle dans toute recherche. En particulier il faut évaluer de manière indépendante, sur plusieurs échantillons, par plusieurs chercheurs et différentes recherches, les propriétés des mesures employées. Enfin, ils argumentent à partir de Cronbach (1971) que la mesure des construits n'est pas une simple validation empirique d'une mesure, mais prend son sens à la lumière des relations théoriques élaborées.

Force est de constater que d'une manière générale les recommandations de Churchill et de Venkatraman et Grant ne sont pas suivies. Dans le cas des recherches cross-culturelles, sous prétexte de différence culturelle, alors que les problèmes de traductions sont ignorés, certains chercheurs s'empressent d'invoquer la différence culturelle, et ignorent les phases de fiabilité et de validité de l'instrument de mesure. De telles conclusions sont difficiles à admettre, car



leurs construits empiriques ainsi mis en évidence ne sont pas les mêmes que les construits issus de recherches antérieures.

Le regroupement d'énoncés qui sont censés rendre compte de manière indirecte du phénomène sous-jacent au construit que le chercheur souhaite appréhender et mesurer est une des étapes de ce que l'on appelle la psychométrie (Nunnally, 1978). La seconde étape de la psychométrie concerne l'évaluation de l'échelle de mesure que le chercheur a créée. Cette évaluation fait appel à des critères liés à la fiabilité et à la validité de l'instrument de mesure (Venkatraman et Grant, 1986).

La fiabilité est une condition nécessaire pour que la validité puisse être établie. Elle se rapporte essentiellement au degré avec lequel un instrument de mesure ne possède pas d'erreur de mesure aléatoire et conduit de cette façon à mesurer effectivement ce que l'on cherche à mesurer. Pour répondre à cette exigence de fiabilité, la méthode la plus couramment utilisée est l'alpha de Cronbach (Fink, 1995). On peut rappeler que l'alpha de Cronbach est une mesure qui repose sur les valeurs de variance et de covariance entre les items. Ce coefficient, issu des travaux en psychométrie sur la mesure de l'erreur, s'interprète comme la corrélation de l'échelle, développée avec tous les autres instruments de mesure du même construit. Dans le cadre d'études internationales, on pourra vérifier si les échelles de mesure possèdent sur chaque dimension le même alpha de Cronbach dans chacun des pays où l'échelle est administrée.

Dans le cadre des études internationales, en complément de la fiabilité, il est aussi important d'évaluer la *dimensionnalité* des échelles de mesure dans chacun des pays où l'étude est effectuée (Venkatraman et Grant, 1986). Si l'échelle est véritablement applicable à l'étranger, sa structure de facteur(s) et le modèle des poids factoriels devraient être équivalents à travers toutes les cultures. Si le modèle des poids factoriels n'est pas cohérent selon les pays où l'on effectue l'étude, alors il est probable que des construits différents sont évalués, et dans ce cas une conclusion basée sur des différences culturelles ne peut pas être valide. Dans la recherche présentée ici, l'analyse factorielle est utilisée pour évaluer la dimensionnalité de l'instrument de mesure dans chacun des pays où le questionnaire est administré.

Plusieurs types de validité peuvent être estimés d'après Peter (1981) : la validité de contenu, la validité convergente, la validité discriminante, et la validité nomologique. Toutefois, dans le cadre de l'évaluation de l'échelle de mesure que le chercheur effectue pour essayer d'appréhender et de mesurer un concept, la validité de construit est l'indicateur le plus saillant



de la validité de mesure et elle est ordinairement scindée en deux aspects : la validité convergente et la validité discriminante qui permettent de vérifier qu'une échelle mesure bien ce qu'elle prétend mesurer (Cf. aussi Venkatraman et Grant, 1986).

A priori, les validités convergente et discriminante se valident au moyen de la procédure multitraits-multiméthodes (Campbell et Fiske, 1959). L'approche liée à la MTMT (MultiTraits-MultiMethods) pour évaluer la validité de construit de Campbell et Fiske a été remplacée récemment par l'Analyse Factorielle Confirmatoire (AFC) (Igalens et Rousel, 1998). Comme le rappelle Peter (1979) et Roehrich (1993), la MTMT n'est pas exempte de critiques et la qualité du résultat n'est pas garantie. L'AFC offre des avantages nombreux, dont le développement de mesures d'ajustement du modèle de mesure général, des informations utiles sur les validités convergente et discriminante, et des méthodes pour déterminer les termes d'erreur des échelles de mesure (Igalens et Roussel, 1998). Cependant, pour avoir le maximum de rigueur dans l'établissement de la validité globale du construit, l'utilisation conjointe de ces deux méthodes est recommandée par Venkatraman et Grant (1986).

A l'aide des considérations méthodologiques ci-dessus, nous avons opéré l'adaptation d'une échelle de mesure nord-américaine d'un domaine stratégique important, l'environnement. Après avoir rapidement brossé son rôle dans les analyses stratégiques, nous allons décrire les étapes d'adaptation du questionnaire, puis décrire les résultats de nos mesures. Nous discuterons pour finir l'intérêt et les limites de ce travail quant à la question méthodologique de l'adaptation des échelles de mesures.

4. LE CONCEPT A MESURER : LA PERCEPTION DE L'ENVIRONNEMENT

Dans une économie de marché qui se veut de plus en plus compétitive et globale, la prise en considération de l'environnement peut être une des variables qui contribuent à la survie et au développement de l'entreprise. La variable liée à l'environnement n'est pas nouvelle dans les recherches en gestion, et elle a donné lieu à de nombreuses controverses (Orpen, 1994).

D'un côté, le courant de l'école de l'écologie des populations part de l'idée que l'environnement est une donnée objective sur lequel l'entreprise ne peut agir. L'environnement sélectionnerait ainsi les entreprises, sans que celles-ci ne puissent véritablement influencer sur leurs chances de succès (Aldrich, 1979)

De l'autre côté, le courant du choix stratégique reconnaît que l'environnement est une réalité perçue par les acteurs. L'environnement ne serait pas une donnée, mais plutôt un construit



(Weick, 1969). Les dirigeants d'entreprise décrètent et mettent en scène l'environnement auquel leur entreprise doit s'adapter à partir des représentations mentales qu'ils se font de la réalité qui les entoure (Silvestre et Goujet, 1996). C'est à partir de ces représentations mentales que les dirigeants d'entreprises déterminent leurs choix stratégiques afin d'atteindre une meilleure cohérence entre l'entreprise et l'environnement dans lequel exerce l'entreprise.

Dans une perspective plus nuancée, le courant de la contingence part de l'idée que les dirigeants d'entreprise structurent leur organisation et organisent leur stratégie en fonction de l'évolution de l'environnement de leur entreprise, et que dans ces conditions, l'adaptation de l'entreprise est un impératif de survie à leur entreprise (Miller et Friesen, 1982, 1983) Dans le même état d'esprit, les auteurs qui se sont intéressés à cette contingence environnementale ont spécifié les dimensions de l'environnement qui déterminaient la relation entre l'environnement et l'organisation, dans l'optique de l'amélioration des performances de l'entreprise.

La contextualisation des stratégies globales et de GRH ne peut se comprendre que si on intègre la manière dont les entreprises perçoivent leur environnement. Les entreprises ne répondront certainement pas de la même manière en fonction des pays où elles sont implantées. Miller (1982, 1983) a montré dans une logique configurationnelle que les entreprises les plus performantes sont celles qui possédaient une harmonisation entre un certain type d'environnement, une certaine stratégie, et une certaine structuration. Un instrument de mesure de l'environnement devrait aider, d'une part, le chercheur à mieux comprendre les relations théoriques entre plusieurs variables, et, d'autre part, les entreprises à analyser les décalages entre leur environnement et leur configuration organisationnelle, dans le cadre de positionnements soit nationaux soit internationaux afin d'y rechercher une meilleure adéquation.

C'est dans cette dernière perspective que D. Miller et P. H. Friesen (1982, 1983) ont étudié l'environnement de l'entreprise à partir des perceptions émises par les chefs d'entreprise. Leurs travaux ont permis de construire une échelle de mesure des perceptions de l'environnement, échelle dont ils ont assuré fiabilité et validité de construit. Ces perceptions liées à l'environnement sont représentées par trois dimensions qui sont celles liées au dynamisme, à l'hétérogénéité, et à l'hostilité de l'environnement auquel est confrontée l'entreprise.



Miller et Friesen expliquent que la première dimension se réfère à l'incertitude liée au changement dans les préférences des clients, dans les technologies de production ou de service, et les modalités concurrentielles dans le principal secteur industriel de l'entreprise. La deuxième dimension, l'hétérogénéité, est liée aux différences entre les différents marchés sur lesquels intervient l'entreprise en ce qui concerne les tactiques concurrentielles, les préférences des clients, les gammes de produit, et les circuits de distribution. Ces différences ne prennent de sens que dans la mesure où elles exigent une commercialisation différente ainsi que des pratiques administratives et de production différente. Enfin, la dernière dimension, l'hostilité, apparaît à travers la concurrence sur le prix, le produit, la technologie et la distribution, via des restrictions réglementaires, des manques de main-d'œuvre ou de matériaux (rationnements), ou encore par des tendances démographiques défavorables (par exemple des marchés décroissants).

Cette échelle de mesure liée au concept de l'environnement, lors des recherches effectuées à son sujet, a prouvé que dans un contexte nord américain, elle était fiable et qu'elle répondait de manière satisfaisante au critère de validité (Merz, Sauber, 1995, Dess et al., 1997). Elle semble être un bon instrument de mesure du construit lié à l'environnement, d'où son utilisation dans la communauté de recherche aux USA. De plus, elle est facile à administrer, car relativement courte.

Malheureusement, comme la plupart des échelles développées outre atlantique, l'échelle de mesure élaborée par Miller et Friesen (1982, 1983) n'a jamais été testée à l'étranger. Sa validité interculturelle et sa fiabilité n'ont pas été étudiées. Nous nous proposons d'évaluer les propriétés psychométriques de cette échelle de mesure dans une logique interculturelle.

5. L'ADAPTATION DU QUESTIONNAIRE

5.1. METHODOLOGIE

L'échelle de mesure utilisée dans ce travail de recherche a été utilisée dans un contexte nord américain de manière assez fréquente (voir annexe 1, le questionnaire en anglais). Pour pouvoir parvenir à élaborer un questionnaire fidèle à la version anglo-saxonne, et ainsi pouvoir l'utiliser dans le cadre du contexte français, nous avons utilisé plusieurs des précautions méthodologiques conseillées par la littérature (Cf. ci-dessus), ce qui explique que la version finalisée en Français du questionnaire a été réalisée un an et demi après avoir commencé à être traduire.



Etapes de la création des items en français

Vallerand (1989) constate qu'il existe trois méthodes pour traduire un questionnaire, appelées méthode traditionnelle, méthode du comité (voir aussi Forgues, 1995, et sa variante de la traduction parallèle aveugle, Usunier, 1992), et méthode de la traduction inversée (développée originairement par Whiting, 1968). Dans la plupart des travaux de recherche, la méthode traditionnelle est la plus utilisée, malgré les nombreuses critiques qu'elle comporte (Douglas, Craig, 1983). Nous avons adopté dans ce travail la méthode de la traduction inversée, qui semble être la méthode la plus lourde à mettre en place, mais aussi très nettement la plus fiable (Ibid). Nous l'avons combinée avec la traduction parallèle aveugle, en faisant traduire de manière indépendante par deux traducteurs, puis en créant un traduction relevant des deux propositions, puis en faisant deux rétro-traductions en aveugle, par deux autres traducteurs. En complément de ce que propose Vallerand (1989), nous avons contacté à plusieurs reprises le professeur Danny Miller afin d'obtenir des précisions sur certaines des expressions qui ont été sources d'écarts lors des différentes traductions et rétro-traductions. Après avoir reçu des précisions de cet auteur sur les expressions qui posaient problème, nous les avons adaptées, et nous les lui avons soumises pour avis. Ce n'est qu'une fois que nous avons reçu ses avis favorables que nous avons fait partir le questionnaire.

D'une manière analytique, la procédure que nous avons utilisée pour générer les items se décompose de la manière suivante.

1/ Traduction parallèle :

L'échelle a été traduite de l'anglais en français par deux professeurs d'anglais de l'université de Lille. Nous avons eu recours à des professeurs d'anglais spécialisés dans le langage du management et disponibles au moment où nous avons besoin d'eux pour traduire le questionnaire. Par la même, afin d'éviter des biais d'influence, nous leur avons demandé de faire le travail seul, car nous avons l'intention de demander à d'autres collègues d'intervenir dans une phase ultérieure.

2/ Traduction inversée parallèle :

L'échelle a été traduite du français à l'anglais par deux autres professeurs d'anglais de l'Université de Lille 1.

3/ Contrôle de traduction :



Une comparaison a été faite par nous même entre la version anglaise du questionnaire initial et la version qui a été traduite puis retraduite. Lorsque des écarts sont apparus, une discussion a eu lieu avec chacun des traducteurs.

4/ Pré-test « acteurs » du questionnaire :

Il s'est déroulé par entretiens en tête-à-tête auprès de 4 chefs d'entreprises et de 2 consultants spécialisés du domaine. Ces entretiens ont lieu de manière séparée avec les chefs d'entreprise, à l'exception des consultants que nous avons rencontré ensemble. Il est ressorti de ces entretiens que les items étaient faciles à comprendre par les personnes interrogées. Pour chacun des items, dans un souci de vérifier son contenu, nous avons demandé à chaque personne interrogée de nous dire ce qu'elle entendait par l'item. Aucun contresens n'est apparu.

Cependant, lors des différents entretiens, il est apparu que l'échelle devait utiliser une mesure qui puisse être facile à comprendre par le répondant. Dans un premier temps, nous avons traduit les termes mêmes qu'avaient établi Miller et Friesen. Mais lors de la phase du pré-test, les répondants nous avons fait part de la lourdeurs des expressions traduits de l'échelle originale (Miller et Friesen proposaient deux phrases antagonistes, entre lesquelles les répondants devaient se situer, Cf. l'annexe 1). Nous avons adapté l'échelle de mesure en tenant compte des recommandations des chefs d'entreprise. Dans l'échelle en français, nous avons choisi une échelle qui s'exprime entre l'accord complet ou le désaccord complet à une proposition. Chacun des items a été évalué sur une échelle à sept points où chaque répondant devait donner son degré d'importance allant de “ pas du tout d'accord ” (1) à “ tout à fait d'accord ” (7).

5/ Pré-test « chercheurs » du questionnaire :

Nous avons échangé avec deux professeurs et trois maîtres de conférences en Sciences de gestion afin d'évaluer la clarté des énoncés et la compréhension des contenus.

6/ Contact direct avec le professeur D. Miller :

Ces contacts nous ont permis d'avoir des précisions sur sa version anglaise du questionnaire à partir des résultats des deux pré-tests ci-dessus. Nous avons pu retoucher quelques détails dans les expressions, et avoir son avis positif sur notre interprétation de son travail. Nous avons pu d'intégrer ses remarques, notamment sur le fait d'utiliser une échelle construite avec une perspective « non restrictive » du construit, même au risque d'avoir des corrélations faibles entre les items d'une même dimension.



Par « non restrictive », il faut entendre le fait de garder des items qui corrélaient plus faiblement avec le construit que d'autres, mais qui ont un sens similaire pour les processus organisationnels. Dans une optique de recherche de la meilleure corrélation (perception restrictive du construit), ces items auraient été éliminés. Dans une perspective plus large, cherchant plus la richesse et la diversité des items contribuant au construit que le meilleur résultat statistique, ces items partiellement redondants sont gardés.

D. Miller, lors d'un échange en 2000 nous a ainsi écrit :

« Which brings me to the environmental scales. They are multifaceted in the sense that a dimension such as uncertainty is composed of several very different things, say technological and competitive uncertainty, and deregulation, that might not be related statistically -- but might nonetheless have similar implications for some of an organization's processes. In that case, even though items may not correlate highly, there is value to having an additive scale that combines them. Of course this is ONLY true if the scale does in fact have predictive value. because of the "substitutability" of the items. »

La dernière remarque de Miller doit être gardée à l'esprit pour de futures utilisations de cette échelle multidimensionnelle. Lors d'une utilisation prédictive (sur d'autres construits), il sera nécessaire de tester les corrélations sur la version non restrictive et sur une version restrictive, avec une épuration plus sévère.

Epuration de la mesure

Comme nous l'avons indiqué plus haut, cet article suit une démarche méthodologique inspirée du paradigme de Churchill (1979). Ce paradigme a été élaboré par cet auteur pour construire des échelles de mesure multiple (dans ces questionnaires à échelle de mesure, on cherche à utiliser plusieurs indicateurs pour mesurer le même construit). Il peut être décomposé en deux grandes phases. La première phase dite « exploratoire » cherche à travers une procédure d'épuration à éliminer et à modifier des items qui ne correspondent pas aux facteurs auxquels ils devraient être affectés. Cette procédure cherche à réduire l'erreur aléatoire à partir des coefficients de cohérence interne qui testent la fiabilité des échelles de mesure. La seconde phase dite phase de « confirmation » cherche à réduire l'erreur systématique en faisant en sorte que la mesure soit spécifique, et qu'elle mesure bien ce qu'elle est censée mesurer. Toutefois, ce cadre a été quelque peu modifié, pour pouvoir prendre en considération les apports des méthodes dites de seconde générations (Analyse Factorielle Confirmatoire), et de



la réalité et des opportunités rencontrées sur le terrain par notre propre processus de recherche.

Sur la base de la matrice inter-items et des analyses factorielles avec rotation oblique¹, nous avons effectué deux pré-tests du questionnaire français auprès de 2 échantillons de 350 et 300 entreprises respectivement, afin d'éliminer ou de modifier certains énoncés (voir annexe 1). Par rapport aux données que nous cherchons à exploiter, les pré-tests ne pouvaient pas être réalisés auprès d'échantillon de convenance, en conséquence ils ont été réalisés auprès d'une population de chefs d'entreprises.

L'analyse factorielle avec rotation oblique a été choisie dans la mesure où comme le montrent Miller et Friesen (1982, 1983) en ce qui concerne la perception de l'environnement, les dimensions qui la composent ne sont pas indépendantes, mais au contraire corrélées. Dans un cas comme celui-ci, et comme le recommande Pedhazur et Pedhazur Schmelkin (1991), il est préférable d'utiliser des rotations obliques plutôt que des rotations orthogonales.

Pour pouvoir épurer la mesure sur l'ensemble des questionnaires revenus, 80 questionnaires ont été retenus pour l'exploitation du premier pré-test, et 75 lors du second. A l'issue de ces pré-tests trois dimensions ont été identifiées, comme dans le questionnaire de Miller et Friesen. Par comparaison aux résultats de Miller et Friesen (1982, 1983), certains items ont dû être réécrits puisqu'ils ne corrélaient pas de manière significative avec leur sous dimensions respectives.

Détermination de l'échantillon

Les questionnaires en langue française et en langue anglaise ont été administrés par voie postale au cours du printemps de l'année 2000, en France pour le premier, et aux Etats-Unis pour le second. Pour établir une homogénéité dans la population que nous cherchions à interroger, seules les entreprises répondant aux critères suivants ont été retenues :

- entreprises de type familial (ont donc été exclues les entreprises filiales de grands groupes ou les filiales de P.M.E.),
- le capital est détenu par des personnes physiques,
- entreprises dont l'effectif est compris entre 5 salariés minimum et 500 salariés maximum.

¹ Pour plus de précision sur les termes techniques, nous renvoyons le lecteur à l'ouvrage de Igalens J., Roussel P. (1998), ainsi qu'à Thiétard et coll. (1999), particulièrement le chapitre 13, C. Donada et A. Mbengue, « Méthodes de classification et de structuration », p.373 sq.



L'échantillon final utilisé en phase de validation comprend 2000 entreprises pour la France, et 800 entreprises pour les Etats-Unis. Nous avons eu un retour de 190 questionnaires exploitables pour la France et 88 pour les Etats-Unis. L'effectif moyen des entreprises américaines était de 12, et de 24 pour les entreprises françaises, avec un écart type de 2 et de 3 respectivement.

Pour ces échantillons, il avait été demandé à ce que ce soit le dirigeant de l'entreprise qui y réponde, d'une part pour s'assurer de l'homogénéité des répondants et d'autre part du fait de la confidentialité et de l'importance des informations que nous cherchions à obtenir. Rappelons que si dans des recherches cross-culturelles, la représentativité des échantillons est importante pour véritablement comparer des populations, on peut estimer que cette représentativité est acceptable du fait de la taille des entreprises qui ont répondu aux questionnaires et du fait que ce soit les chefs d'entreprise qui aient renseigné le questionnaire, même si comme le reconnaissent Green et White (1976), arriver à une parfaite comparabilité est très difficile, voire même impossible sur le plan pratique (Cf. la littérature relevée ci-dessus).

La taille et les caractéristiques de l'échantillon utilisé pour la validation de l'échelle de mesure peuvent donc être considérés comme pertinent pour notre l'objet de cette étude. Certaines limites sont abordées dans la discussion de nos résultats.

Evaluation de la fiabilité et de la validité de mesure

Pour traiter les réponses du questionnaire final en français et du questionnaire en anglais et pour répondre aux hypothèses de recherches que nous avons formulées, des analyses en composantes principales avec rotation oblique ont été réalisées sur l'ensemble des deux questionnaires. Après avoir établi la dimensionnalité de l'instrument de mesure, nous en avons vérifié la fiabilité à partir de l'alpha de Cronbach.

Suite à ce travail un modèle qui pose *a priori* des relations entre les différentes dimensions du concept et les indicateurs respectifs à chaque dimension a été réalisé à partir des résultats de l'analyse factorielle. Ce modèle a été testé en effectuant une analyse factorielle confirmatoire aussi bien pour les réponses obtenues à partir du questionnaire français que du questionnaire anglais. La qualité des résultats a été testée à partir de LISREL VII, en utilisant les indices suivants qui sont ceux qui sont utilisés de manière courante:

- le *goodness of fit Index* (GFI) et l'*adjusted goodness of fit Index* (AGFI) indiquent la part de la variance expliquée par le modèle.



- le RMR (*root mean square residuals*) qui représente la valeur moyenne des résidus.
- le χ^2 rapporté au nombre de degrés de liberté, qui est un critère de bon ajustement du modèle ; plus il est proche de 0 meilleur est le modèle. Toutefois, il est utile de rappeler que cet indice même rapporté au nombre de degrés de liberté demeure sujet aux variations de taille de l'échantillon.
- Enfin, pour être certain que le modèle français et américain qui suppose l'existence de trois dimensions possède une supériorité par rapport à d'autres modèles, plusieurs modèles de dimensions ont été comparés au modèle nul qui suppose l'absence d'une structure factorielle.

Une fois que la qualité de l'ajustement a été établie, nous avons cherché à établir la validité convergente et la validité discriminante des échelles de mesure.

Dans le cadre d'échelles multidimensionnelles, la validité discriminante est vérifiée à condition que la variance partagée entre les construits soit inférieure à celle partagée entre les construits et leurs variables de mesure. Pour le cas de la validité discriminante, et dans le cadre d'échelles multidimensionnelles, elle est satisfaisante lorsque la variation du χ^2 est supérieure à 6.64 au seuil de 1% pour un degré de liberté. Elle a été calculée à partir des variations du χ^2 obtenues en comparant le χ^2 entre le modèle contraint (en fixant à 1 le paramètre de corrélation entre 2 construits) et le modèle non contraint, ou le paramètre de corrélation est laissé libre (Bagozzy et Philipps, 1982). En ce qui concerne la validité convergente, nous avons retenu la méthode préconisée par Larcker et Formell (1981) qui devient courante en recherche. Pour cette méthode, la validité est établie lorsque la variance moyenne partagée entre les variables de mesure et la variable latente qui lui est rattachée est supérieure à 50%.

5.2. HYPOTHESES DE RECHERCHE

La discussion précédente nous mène aux hypothèses de recherche suivantes sur l'échelle de mesure du concept lié à l'environnement de la firme :

H1: Le questionnaire traduit en français a une structure factorielle en trois dimensions comme pour le questionnaire issu de l'anglais.

H2: La structure des saturations au niveau des poids factoriels est invariante entre le questionnaire anglais et le questionnaire traduit en français.

H3: Les coefficients de cohérence interne (alpha de Cronbach) français sont élevés, et ils sont équivalents entre le questionnaire anglais et le questionnaire français.

H4 : Les questionnaires en anglais et en français démontrent une validité convergente pour mesurer chacune des dimensions qui sont liées au concept lié à l'environnement.

H5 : Les questionnaires en anglais et en français démontrent une validité discriminante pour mesurer chacune des dimensions qui sont liées au concept lié à l'environnement.

6. RESULTATS

Avant de procéder à un travail lié à l'analyse factorielle confirmatoire, une démarche exploratoire a donc été utilisée sur la base de l'analyse factorielle exploratoire et sur la base de la matrice des corrélations inter-items pour chaque questionnaire. Comme attendu, le concept lié à la perception de l'environnement aussi bien pour l'étude réalisée en France que pour l'étude réalisée aux Etats-unis est un construit multidimensionnel (tableau 1). Ces résultats corroborent ceux qui avaient été établis précédemment par Miller et Friesen (1982, 1983). D'une manière plus détaillée, chacun des items est corrélé avec son facteur respectif, à l'exception des items 1, 2 et 11 qui ont été supprimés (faute d'une corrélation forte avec leur facteur respectif au niveau de l'échelle d'origine). Ainsi, les items 3 à 5 sont corrélés avec la dimension liée au dynamisme, les items 9, 10 et 12 sont corrélés à la dimension liée à l'hétérogénéité, et enfin les items 6, 7 et 8 sont corrélés à la dimension liée à l'hostilité.

Tableau 1 : Résultats de l'analyse factorielle exploratoire

Questionnaire	Anglais			français		
	DY	HO	HE	DY	HO	HE
Items						
3 -Le rythme de l'innovation dans les procédés de production s'est accéléré considérablement	0.89			0.86		
4 -Le rythme de l'innovation en nouveaux produits et services s'est accéléré considérablement	0.90			0.86		
5 - Les activités de recherche et de développement ont augmenté substantiellement	0.83			0.81		
6 - Les démarches commerciales de vos principaux concurrents affectent votre entreprise						
7 - Les démarches commerciales de vos principaux concurrents peuvent être anticipées plus facilement		0.76			0.72	
8 - Les démarches commerciales de vos principaux concurrents sont devenues moins agressives		0.92			0.78	
		0.71			0.91	
9 - Une plus grande diversité des productions/services est nécessaire afin de les adapter à vos clients			0.65			0.65
10 - Une plus grande diversité des prod/services est nécessaire afin de l'adapter au dynamisme marché			0.89			0.86
12 - Une plus grande diversité des moyens de prod/ gestion est nécessaire afin de répondre à la concurrence			0.93			0.89
Coefficient de Cronbach	0.86	0.76	0.74	0.81	0.75	0.87

Dy = dynamisme, Ho = hostilité, He= Hétérogénéité

Afin de s'assurer de l'adéquation du modèle de mesure spécifié avec les données empiriques sous-jacentes, nous avons utilisé plusieurs critères d'ajustement (Cf. ci-dessus). Les résultats des analyses factorielles exploratoires sont confirmés par les analyses factorielles confirmatoires puisque les indicateurs d'ajustement du modèle sont bons autant pour le questionnaire en anglais (GFI=0.918, AGFI=0.846, RMR=0.069, $\chi^2=32.14$ $\chi^2/ddl=1.34$) que pour le questionnaire en français (GFI=0.931, AGFI=0.872, RMR=0.069, $\chi^2=68.10$, $\chi^2/ddl=2.83$). En conséquence, on peut en conclure que le modèle qui établit l'existence de trois dimensions pour mesurer l'environnement de l'entreprise ajuste convenablement les données. Toutefois, plutôt que de prendre seulement en compte la qualité absolue de l'ajustement de ce modèle qui établit l'existence d'une structure en 3 dimensions, nous avons comparé plusieurs modèles au modèle nul.

Ainsi, le modèle qui suppose l'existence d'un seul facteur, de deux facteurs, et de 3 facteurs ont été comparés au modèle nul (Tableau 2). L'analyse se faisant par la comparaison des différents modèles au modèle nul, il ressort que le modèle conceptuel qui établit l'existence de 3 dimensions a le meilleur ajustement aux données empiriques que ce soit pour le questionnaire en français ou en anglais. A partir de ces résultats, nous pouvons dire que l'hypothèse 1 : « *Le questionnaire issu de l'Anglais a une structure factorielle en trois dimensions comme pour le questionnaire traduit en français* » est vérifiée.

Tableau 2 Comparaison des modèles d'analyses factorielles confirmatoires

Indices d'ajustement	Modèle nul :	Modèle ayant 1 dimension	Modèle ayant 2 dimensions	Modèle ayant 3 dimensions
Questionnaire français				
χ^2	229.28	105.20	56.58	33.06
dl	36	27	26	24
χ^2/dl	6.36	3.88	2.19	1.38
GFI	0.604	0.794	0.884	0.917
AGFI	0.256	0.657	0.800	0.845
RMR	0.321	0.136	0.102	0.069
Questionnaire anglais				
χ^2	402	154.16	112.49	32.14
dl	36	27	26	24
χ^2/dl	11.16	5.71	4.33	1.34
GFI	0.449	0.725	0.798	0.918
AGFI	0.311	0.541	0.650	0.846
RMR	0.351	0.167	0.164	0.069



De même, l'hypothèse 2 : « La structure des saturations au niveau des poids factoriels est invariante entre le questionnaire en anglais et le questionnaire traduit en français » est vérifiée. Deux modèles sont construits et estimés pour effectuer cette analyse. Pour le premier modèle, tous les paramètres sont laissés libres. Pour le second modèle, une contrainte d'égalité est imposée entre le poids factoriel des items de chaque questionnaire, respectivement en français et en anglais. Le χ^2 est de 67.83 (ddl =48) pour le premier modèle, et de 88.92 (ddl =57) pour le second. La différence entre ces 2 valeurs (χ^2 de 21.09, ddl=9) n'est pas significative au seuil de 5% permet de conclure que la structure factorielle est invariante entre la France et les Etats-Unis (Joreskog et Sorbom, 1989), et que l'on possède une équivalence des instruments de mesure.

Nous avons établi la dimensionnalité de chaque dimension, en vérifiant que les intensités qui lient une dimension (variable latente) à ses indicateurs de mesure sont significativement différentes de 0, c'est-à-dire lorsque la valeur de t est supérieure à 1.96. Nous avons vérifié la fiabilité des différentes échelles qui composent le questionnaire en utilisant l'alpha de Cronbach. Les résultats présentés dans le tableau 3 montrent que les alphas de Cronbach sont tous supérieurs à 0.7, ce qui indique que les items mesurent respectivement les dimensions qu'ils sont censés mesurer. En outre, à la lecture des alphas de Cronbach du tableau 3, il ressort qu'il n'y a pas de différence majeure entre les alphas de Cronbach de la version anglaise et de la version française pour chacune des dimensions (dynamisme, hostilité, hétérogénéité).

De ce fait on peut dire que les différentes dimensions du concept lié à l'environnement, que ce soit pour la version française ou anglaise du questionnaire, mesurent de manière équivalente leur construit. Les résultats de ce tableau font dire que l'hypothèse n°3 : « *Les coefficients de cohérence internes français sont élevés et équivalents entre le questionnaire anglais et le questionnaire français* » est vérifiée.

Tableau 3 Résultats des contributions de l'analyse factorielle confirmatoire

Dimensions	Anglais			Français		
	Dy	Ho	He	Dy	Ho	HE
3 -Le rythme de l'innovation dans les procédés de production s'est accéléré considérablement	.84 (11.2)			.82 (11.6)		
4 -Le rythme de l'innovation en nouveaux produits et services s'est accéléré considérablement	.92 (12.9)			.71 (9.9)		
5 - Les activités de recherche et de développement ont augmenté substantiellement	.71 (8.9)			.77 (10.9)		
6 - Les démarches commerciales de vos principaux concurrents affectent votre entreprise		.66 (7.4)			.61 (7.9)	
7 - Les démarches commerciales de vos principaux concurrents peuvent être anticipées plus facilement		.81 (9.3)			.75 (9.6)	
8 - Les démarches commerciales de vos principaux concurrents sont devenues moins agressives		.72 (8.2)			.80 (10.4)	
10 - Une plus grande diversité des prod/services est nécessaire afin de l'adapter au dynamisme marché			.72 (7.0)			.53 (4.6)
12 - Une plus grande diversité des moyens de prod/ gestion est nécessaire afin de répondre à la concurrence			.99 (18)			.87 (5.4)

Les valeurs de t figurent entre parenthèses

En ce qui concerne le teste de l'hypothèse 4, le tableau 3 montre que les estimations des contributions de chacun des items sur leur construit respectif sont corrélées de manière significative, et que la variance capturée moyenne de chaque construit pour le questionnaire en anglais est de 0.65 pour le construit lié au dynamisme, de 0.54 pour le construit lié à l'hostilité, et de 0.75 pour le construit lié à l'hétérogénéité. Pour le questionnaire en français, on peut également constater dans le tableau 3 que les estimations des contributions de chacun des items sur son construit sont significativement corrélées. La variance capturée moyenne de chaque construit pour le questionnaire en français est de 0.58 pour le construit lié au dynamisme, elle est de 0.53 pour le construit lié à l'hostilité, et de 0.52 pour le construit lié à l'hétérogénéité. En fonction de ce résultat, et si on applique la méthode de Larcker et Formell (1981), on peut dire que la validité convergente est satisfaisante pour tous les construit en anglais, et qu'elle l'est aussi pour les construits en français. Au vu de ces résultats, on peut estimer que l'hypothèse 4 : «*Les questionnaires en anglais et en français démontrent une validité convergente pour mesurer chacune des dimensions qui sont liées au concept lié à l'environnement* » est vérifiée.



Enfin pour l'hypothèse 5 : « Les questionnaires en anglais et en français démontrent une validité discriminante pour mesurer chacune des dimensions qui sont liées au concept lié à l'environnement », la validité discriminante a été vérifiée pour tous les construits. Comme le préconisent Bagozzi et Yi (1988), la validité discriminante peut être vérifiée en comparant le χ^2 obtenu entre le modèle contraint (en fixant à 1 le paramètre de corrélation entre deux construits) et le modèle non contraint (ou le paramètre de corrélation est laissé libre). En utilisant LISREL, la validité discriminante a pu être déterminée en comparant la différence des χ^2 entre ceux obtenus du modèle où tous les paramètres de corrélation entre les construits sont laissés libres (24 degrés de liberté) et ceux obtenus du modèle où l'on contraint le paramètre de corrélation à 1 pour les construits pour lesquels on cherche à vérifier la validité discriminante (25 degrés de liberté). Comme l'indique le tableau 5, la validité discriminante est établie pour tous les construits dans la mesure où la différence des χ^2 obtenus est significative au seuil de 1%, aussi bien pour le questionnaire en Français que pour le questionnaire en anglais.

Tableau 5 : Vérification de la validité discriminante

Les construits ou le paramètre De corrélation est fixé à 1	différence du χ^2 pour 1 degré de liberté pour le questionnaire français	Significatif	différence du χ^2 pour 1 degré de liberté pour le questionnaire anglais	Significatif
Dynamisme/ hétérogénéité	13.71	P<0.001	29.08	P<0.001
Dynamisme/ Hostilité	12.48	P<0.001	19.20	P<0.001
Hétérogénéité / Hostilité	21.83	P<0.001	26.31	P<0.001

7. DISCUSSION, CONCLUSIONS, PERSPECTIVES ET LIMITES

Le premier apport de cette recherche se situe au niveau de la démarche elle-même. La conception d'un instrument de mesure d'origine américaine pour l'adapter au contexte français avec un bon degré d'équivalence n'est pas une chose facile. En effet, à partir des conseils dégagés de la littérature relevant d'une approche hybride, il faut du temps et de la rigueur tant dans la démarche pratique que dans la réflexion intellectuelle. Il faut du temps, car entre la première version du questionnaire et la dernière, il s'est écoulé plus d'un an et demi, temps nécessaire à la construction du questionnaire, avec plusieurs phases de pré-tests. Ce temps, même s'il paraît long, a permis d'échanger tant avec des chefs d'entreprises, des consultants, des professeurs de gestion, qu'avec Danny Miller auteur du questionnaire américain que nous avons traduit. Nous avons ainsi enrichi en permanence le travail durant sa



réalisation. Ces échanges ont permis d'épurer l'échelle de mesure et ont permis de s'assurer par la diversité des acteurs que nous avons rencontrés que la validité de contenu a été atteinte.

Le second apport de cette recherche se situe au niveau des résultats. Au terme de cette recherche, il a été possible de s'assurer des qualités psychométriques des instruments de mesure. Il peut dorénavant être utilisé par les chercheurs pour utiliser les dimensions liées au dynamisme, à l'hétérogénéité, et à l'hostilité de l'environnement stratégique, dans leurs études nationales comme dans leurs études comparées entre la France et les Etats Unis, puisque les deux questionnaires semblent équivalents et satisfaire aux critères de fiabilité et de validité à la fois convergente et discriminante. Nous ne commentons ni ne discutons ici ce résultat, qui n'est pas dans notre objet. Notre démarche ne s'est pas intéressé aux pratiques organisationnelles, mais au concept en lui-même. Comme le signalent Gerbin et Anderson (1988), si des indicateurs ont été construits pour mesurer un instrument, lors de la traduction de l'échelle ils doivent refléter le même concept ; dans le cas contraire, on serait loin du concept qui est censé être représenté par une telle échelle de mesure. Ce n'est qu'ultérieurement à cette phase que l'on peut utiliser ce concept pour appréhender la manière dont les individus perçoivent l'environnement de leur entreprise. Nos futurs travaux ne manqueront pas d'utiliser cette échelle de mesure, et de contribuer grâce à elle à l'analyse de plusieurs pratiques organisationnelles.

L'échelle de mesure Nord américaine lorsqu'elle est adaptée en français correspond bien à l'échelle d'origine. Notre travail dépasse le simple exercice de traduction puisque à plusieurs reprises des adaptations ont été réalisées afin de se rapprocher autant que possible du construit établi aux Etats-Unis. La voie choisie, que la littérature permet de dénommer hybride, ou encore intermédiaire, dans le domaine de l'analyse inter-culturel, s'avère avoir donné un résultat significatif. C'est un exercice qui a pris du temps et qui a demandé à s'assurer en permanence d'une méthodologie rigoureuse et qui s'est adaptée aux situations que nous rencontrons. Nous avons utilisé, à chaque étape, les conseils que les travaux de même nature ont prônés. Ainsi, s'il est vrai que les problèmes de traduction ont été nombreux, notre méthodologie ne s'est pas limitée à une rétro-traduction en parallèle (deux rétro-traductions). Sur plusieurs items nous avons dû refaire les échelles en Français car toutes les nuances conceptuelles ne pouvaient pas être transcrites d'une langue à l'autre de manière équivalente par de simples ajustements linguistiques.

Cette échelle de mesure, outre ses bonnes propriétés psychométriques de mesure de chacune des dimensions du concept lié à la perception de l'environnement, a l'avantage d'être courte



et facile à administrer. Sur le plan des implications managériales, utilisée auprès des dirigeants des pays anglophones et des pays francophones lors de recherches internationales ou nationales, elle pourrait servir de base à l'analyse de l'environnement dans lequel les entreprises exercent, champ difficile à cerner, et qui nécessite de plus en plus de posséder un outil méthodologique validé, plutôt que de se fier à une quelconque intuition.

Le troisième apport de cette recherche se veut pédagogique et méthodologique. Si, dans le domaine du marketing, la démarche préconisée par Churchill (1979) semble être utilisée régulièrement, en revanche dans d'autres domaines de recherche, particulièrement en GRH, cette-ci n'est pas encore appliquée de manière systématique (Igalens et Roussel, 1998). Il apparaît qu'une démarche de nature quantitative peut avoir beaucoup d'intérêt lorsqu'elle est assise sur une démarche rigoureuse, mais elle peut également être presque inutile lorsqu'elle manque de rigueur, puisque, dans le cas qui nous intéresse ici, les variables peuvent être mal traduites, et donc mal mesurées.

Concernant les voies de recherches possibles, ce travail ouvre de nouvelles perspectives. Après avoir élaboré une échelle de mesure qui semble satisfaisante sur l'ensemble des sous dimensions, il est envisageable d'entrer dans une perspective de validité prédictive. Concevoir une recherche permettant de montrer l'influence des construits liés à l'environnement sur les stratégies des entreprises, ainsi que sur leurs choix de GRH, est une perspective de recherche à venir.

De manière pragmatique, cette recherche présente trois limites. La première est que pour le moment, le questionnaire a été validé en Français dans un contexte national français. D'autres perspectives sont possibles. Qu'en est-il pour les pays francophones qui ne possèdent pas la même culture que la nôtre ? Nous ne sommes pas certains que ce questionnaire puisse avoir les mêmes valeurs psychométriques en Belgique ou en Suisse francophones, ou au Québec, par exemple.

La seconde limite vient de ce que, pour l'instant, ce questionnaire a été traduit de l'anglais au français, ce qui limite fortement les pays où il peut être administré. En conséquence, un travail ultérieur pourrait être effectué pour traduire ce questionnaire dans d'autres langues, et ainsi permettre des comparaisons qui puissent revêtir un caractère plus largement international.

Enfin, la troisième limite interpelle la bonne qualité de l'échelle en français.

Plusieurs raisons expliquent que cette échelle retranscrite corresponde assez bien à l'échelle américaine. La première vient de ce que nous n'avons interrogé que des chefs d'entreprises.



L'homogénéité des échantillons ainsi créée peut faire perdre une différence culturelle que nous aurions pu retrouver dans d'autres types de population, par exemple des consommateurs, plus sensible peut-être à la différence culturelle que le chef d'entreprise. L'homogénéité se retrouve aussi dans le type des entreprises. Notre échantillon n'était constitué que de petites et moyennes entreprises. Il peut être opposé que ce sont des entreprises qui perçoivent l'environnement d'une manière comparable, que ce soit aux Etats-Unis ou en France. Rien ne garantit en effet que les dirigeants de grandes entreprises approchent ces construits de la même manière en France et aux Etats-Unis.

Enfin, ce résultat ne peut être généralisé à d'autres échelles de mesure que celle qui a été élaborée. Si l'échelle de mesure en français se rapproche bien de l'échelle américaine sur le concept lié à l'environnement, rien ne nous permet de dire que le même échantillon réponde de la même manière sur des variables autres que celle que nous avons convoquées. Qu'en est-il des variables liés au processus de décisions où une connotation managériale est ancrée dans des aspects de pouvoirs et d'autorités très centralisés ? Qu'en est-il des variables stratégiques de positionnement, où semble-t-il les Américains rechercheraient des prises de risques plus importantes que les notre dans la mesure où la notion d'échec n'est pas perçue de manière identique en France et aux USA (A la suite des travaux de Hofstede, la littérature est abondante sur ce point, plus peut-être sur des études de cas que par des approches quantitatives rigoureuses).

Si l'instrument de mesure semble être exploitable dans deux pays, ne serait-ce pas que le concept lié à l'environnement soit une conception américaine dont la conception française se rapprocherait assez bien, du fait de l'imprégnation par le management américain tant des dirigeants français que des chercheurs français en Sciences de gestion ? A titre d'illustration, le lecteur peut se référer au nombre de citations américaines de notre bibliographie. Certaines recherches ont considéré que le système de management est assez proche entre les USA et la France et qu'il relève d'un système de valeur individualiste (Janssens, Brette et Smith 1995) Qu'en est-il des pays où cette conception des choses est éloignée de ces deux pays, et dans une plus large mesure des pays industrialisés ? Qu'en est-il des pays du tiers monde qui vivent des frustrations et des *challenges* tout à fait différents en termes de perspective de ceux des pays riches ? Voient-ils l'environnement économique de la même façon ?

Autant de questions qui nous amènent à n'accorder qu'une légitimité relative et contextuelle à ce premier travail. N'est-il pas temps de concevoir que les Sciences de gestion ne doivent pas uniquement être conçues comme une discipline qui trouve son essence dans les pays



occidentaux, associés à l'efficacité économique, politique et sociale contemporaine, et que même les pays « pauvres » ont « droit » à une conception de l'environnement fondée sur d'autres construits que ceux que nous venons d'élaborer. En conséquence, si dans cette recherche un travail d'ordre quantitatif a suffi à créer une équivalence dans l'instrument de mesure, il se peut qu'un travail d'inspiration anthropologique doive être conçu pour d'autres sociétés afin de véritablement faire spécifier le concept sur lequel nous avons fondé cette recherche, et qui démontrerait le caractère non universel de l'échelle de mesure que nous avons élaboré en français. Les difficultés et les traitements statistiques particulièrement éprouvants (pour le lecteur, mais aussi pour le construit traité) de l'étude ambitieuse de Peterson, Smith, et-al. (1995) sur des échantillons provenant de 21 pays le montrent.

Cette réflexion relève de l'ethnostatistique (Gephart, 1978, 1988). Quelle est la pratique « quotidienne » des chercheurs qui utilisent les méthodologies quantitatives ? Pourquoi utilisent-ils ces outils, quelles méthodologies sont convoquées et pourquoi ? La sophistication des outils méthodologiques statistiques est-elle toujours le garant d'une qualité scientifique ? Le courant critique, voire post-moderne, prône le retour à des techniques non paramétriques, par exemple, censées être plus proches des perceptions et des discours des acteurs (Gephart, 1988, p.33 sq). Plus profondément, quels sont, et pourquoi, les concepts et les construits approchés ? N'y-a-t-il pas un effet de « réverbère », consistant à mener des études seulement sur les concepts et les échantillons qui permettent d'utiliser les outils quantitatifs et permettent leur raffinement ? Autrement dit, les recherches en Sciences de gestion sont-elles guidées par le souci de mieux connaître les phénomènes, en interactions avec les préoccupations des acteurs en organisations, ou par le positionnement des chercheurs dans des champs balisés et reconnus par les pairs ? Cette réflexion comporte ainsi non seulement une grave question épistémologique, mais aussi une dimension proprement éthique.

De ces premiers résultats, il apparaît qu'il existe des échelles de mesure ou la traduction suffit, et ce en cohérence avec l'étude de Davis et al (1981) ainsi que de la revue de littérature en recherche interculturelle de Earley et Singh (1995). La nature des construits, voire des items eux-mêmes, permet parfois d'en rester à des précautions de traduction pour obtenir une bonne équivalence (Usunier, 1992). Mais ce travail contribue *a contrario* à faire garder à l'esprit qu'il existe des échelles de mesure pour lesquelles des adaptations sont nécessaires, et que pour d'autres construits la construction d'une échelle de mesure nouvelle fortement ancrée dans la culture locale s'impose. Notre travail tend à montrer que ce choix n'est pas aussi simple que l'on voudrait le croire. Il dépend de plusieurs facteurs. Il y a en effet, dans



tous les domaines de sciences de gestion, des éléments de pratiques organisationnelles qui ne sont pas, ou peu, liés à la culture et où la notion d'échelle de mesure adaptée semble avoir du sens. Mais il existe aussi des pratiques pour lesquelles l'imprégnation culturelle est tellement forte que chaque pays développe ses propres pratiques très spécifiques, et pour lesquelles il est besoin d'instruments spécifiques (Earley et Singh, 1995).

C'est à partir de l'ensemble des réflexions sur cette appropriation de l'objet (construit) que l'on se rapprochera de ce que l'on veut véritablement mesurer en liant à la fois des échelles de mesures universelles et des échelles adaptées à chaque contexte. Il serait réducteur de croire que tout est globalisé ou que tout est culturellement contextualisé. La complexité des phénomènes que nous avons observés, imbriqués dans des niveaux de conceptualisation de natures différentes, imbrication qui échappe souvent aux analyses ciblées sur des construits bien délimités, nous amène à penser le contraire.

8. RÉFÉRENCES

- Aldrich, H., *Organizations and environment*, Englewood, Cliffs, NJ: Prince-hall, 1979.
- Bagozzi P.P., Yi Y., (1988), On the Evaluation of Structural Equation Models, *Journal of the Academy of Marketing Science*, 16, p. 74-94.
- Churchill G.A., « A Paradigm for Developing Better Measures of Marketing Constructs », *Journal of Marketing Research*, 16, p. 64-73, 1979.
- Davis H.L., Douglas S.P., Silk A.J., (1981), Measure Unreliability : A Hidden Threat to Cross-National Marketing Research ?, *Journal of Marketing*, Spring 1981, p. 98-109.
- Dess G.G., Lumpkin G.T., Covin J.G., (1997), Entrepreneurial Strategy Making and Firm Performance: Tests of Contingency and Configurational Models, *Strategic Management Journal*, Vol. 18, 9, 677-695.
- Divers auteurs (G. Hofstede, Triandis, P. B. Smith), (2001), Special Commentary. The study of Cross Cultural Management and Organization- Past, Present and Future, *International Journal of Cross Cultural Management*, 1, No1, p. 11-30.
- Douglas S.P., Craig C.S., (1984), Establishing equivalence in Comparative Consumer Research, in *Comparative Marketing Systems*, E. Kaynak, R. Savitt, Ed., Praeger, New-York, 93-113.
- Douglas S.P., Craig C.S., *International Marketing Research*, Englewood Cliffs, N-J, Prentice-Hall, 1983.



- Earley P.C., Singh H., (1995), International and Intercultural Management Research : What's next ?, *Academy of Management Journal*, 38, No.2, p. 327-340.
- Evrard Y., Pras B., Roux E., et al., *Market : Etude et recherches en marketing*, Nathan, Paris, 1993.
- Fink Arlène, *The Survey Handbook (The Survey Kit*, vol. 1), et *How to analyse Survey Data (The Survey Kit*, vol. 8), Sage, London, 1995.
- Forgues B., (1995), La traduction des questionnaires, *Nouvelles du Marketing*, Association Française du Marketing, N°34.
- Forgues B., Leconte P., (2000), Compétences des équipes dirigeantes et pratiques de gestion des compétences dans les firmes multinationales,, *Revue française de gestion*, N°127, p.119-130.
- Gephart R.P., Jr., (1978), Status Degradation and Organizational Succession : an Ethnomethodological Approach, *Administrative Science Quarterly*, 23, p. 553-581.
- Gephart R.P., Jr., (1988), *Ethnostatistics : Qualitative Foundations for Quantitative Research*, Sage, London.
- Gerbing D., Anderson J., (1988), An Updated Paradigm for Scale Development Incorporating Unidimensionality and its Assessment, *Journal of Marketing Research*, 25, p. 186-192.
- Hinkin T.R., (1998), A Brief Tutorial on the Development of Measures for Use in Survey Questionnaires, *Organizational Research Methods*, 1, p.104-121.
- Hofstede G., (1987), Relativité culturelle des pratiques et théories de l'organisation), *Revue Française de Gestion*, septembre-octobre.
- Hofstede G., *Vivre dans un monde multiculturel, comprendre nos programmations mentales*, Ed. d'Organisation, Paris, 1994.
- Igalens J., Roussel P. *Méthodes de recherche en gestion des ressources humaines*, Economica, Paris, 1998.
- Iribarne P. D' (1987), Ce qui est universel et ce qui ne l'est pas, *Revue Française de Gestion*, septembre-octobre, p. 6-9.
- Iribarne P. D', *La logique de l'honneur*, Le Seuil, Paris, 1989.
- Janssens M., Brett J.M. , Smith F.J. (1995), Confirmatory Cross-Cultural Research : Testing the Viability of a Corporation-Wide Safety Policy, *Academy of Management Journal*, 38, No.2, p. 364-382.
- Levitt T., (1983), The Globalization of Markets, *Harvard Business Review*, May-June, p. 92-102.



- Lindquist J.D., Vida I., Plank R.E., Fairhurst A., (2001), The modified CETSCALE : validity tests in the Czech Republic, Hungary and Poland, *International Business Review*, Vol.10, 505-516.
- G.R. Merz et M.H. Sauber, (1995), Files of Managerial Activities in Small Firms, *Strategic Management Journal*, 16, p 551-564.
- Miller D., Friesen P., (1982), Innovation in Conservative and Entrepreneurial Firms: Two Models of Strategic Momentum, *Strategic Management Journal*, 3, p. 1-25.
- Miller D., Friesen P., (1983), Strategy-Making and Environment: The Third Link, *Strategic Management Journal*, 4, p. 221-235.
- Netemeyer R.G., Durvasula S., Lichtenstein D.R., (1991), A Cross-National Assessment of the Reliability and Validity of the CETSCALE, *Journal of Marketing Research*, Vol. XXVIII (August 1991), 320-327.
- Nunnally J.C., *Psychometric Theory*, 2nd ed., Mac-Graw-Hill, New York, 1978.
- Nyeck S., Paradis S., Xuereb J.M., (1996), Standardisation ou adaptation des échelles de mesure à travers différents contextes nationaux : L'exemple d'une échelle de mesure de l'innovativité, *Recherche et Application Marketing*, 11, p.57-74.
- Orpen, C. (1994), Strategic Planning, Scanning Activities and the Financial Performance of Small Firms, *Journal of Strategic Change*, 3, p. 45-55.
- Parameswaran R., Yaprak A., (1987), A Cross-National Comparison of Consumer Research Measures », *Journal of International Business Studies*, Spring, p. 35-49.
- Pedhazur E.J., Pedhazur-Schmelkin L., *Measurement, Design and Analysis, An Integrated Approach*, Hillsdale, N-J : Lawrence Erlbaum Associates Publishers.
- Peter T., (1981), Construct Validity : A review of Basic Issues and Marketing Practices, *Journal of Marketing Research*, 18, p. 133-145.
- Peterson M.F., Smith P.B., et al., (1995) Role conflict, Ambiguity, and Overload : A 21-Nation Study, *Academy of Management Journal*, 38, No.2, p. 327-340.
- Robinson, R.B., Pearce A., (1984), Research Thrusts in Small Firm Strategic Planning, *Academy of Management Review*, 9, p. 128-137.
- Schein E., (1990), Organizational Culture, *American Psychologist*, N°45.
- Shimp T.A., Sharma S., (1987), Consumer Ethnocentrism : Construction and Validation of the CETSCALE, *Journal of Marketing Research*, 24, p. 280-289.
- Silvestre, H., Goujet, R., (1996), « Lisibilité de l'environnement, management stratégique : éléments de recherche sur les PMI », *Revue Internationale des PME*. 9, N°1, p. 61-78.



- Smith C.S., Tisak J., Bauman T., Green E., (1991), Psychometric Equivalence of a Translated Circadian Rhythm Questionnaire : Implications for Between- and Within-Population Assessments, *Journal of Applied Psychology*, Vol.76, No. 5, 628-636.
- Tayeb M. (2001), Conducting Research Across Cultures : Overcoming Drawbacks and Obstacles, *International Journal of Cross Cultural Management*, 1, No1, p. 91-108.
- Tayeb M. (2001), Conducting Research Across Cultures : Overcoming Drawbacks and Obstacles, *International Journal of Cross Cultural Management*, 1, No1, p. 91-108.
- Thiétart et coll., *Méthodes de recherche en Management*, Dunod, Paris, 1999.
- Urien B., (2000), Validation d'une Echelle de Mesure d'une Nouvelle Dimension Temporelle en Comportement du Consommateur : l'Anxiété face à l'Avenir, Communication au colloque Le Tendenze del Marketing in Europa, Università Ca' Foscari, Venezia, Novembre.
- Usunier J-C., (1990), Méthodologie de la recherche Marketing Cross-Culturelle : les Limites de l'Equivalence, Actes 1990 de l'association Française du Marketing, La Baules, 10 et 11 mai 1990.
- Usunier J-C., (1992), *Commerce entre cultures, une approche culturelle du marketing international*, Paris, Presses Universitaires de France, collection gestion.
- Valette-Florence P., (1988), Spécificités et apports des méthodes d'analyse multivariée de la deuxième génération, *Recherche et Applications en Marketing*, 3, N°4, p. 23-56.
- Vallerand R.J., (1989), Vers une méthodologie de validation transculturelle de questionnaires psychologiques : implications pour la recherche en langue française, *Canadian Psychology/ Psychologie Canadienne*, 30, N°4, p. 662-680.
- Venkatraman N., Grant J.H., (1986), Construct Measurement in Organizational Strategy Research : A Critique and Proposal, *Academy of Management Review*, Vol.11, No.1, 71-87.
- Weick, K.E. (1969), *The social psychological of organizing*, Reading, MA: Addison-Wesley.
- Whiting (1968) : Whiting J.W.M. (1968), Methods and Problems in Cross-Cultural Research. In G.Lindzey & E.Aranson, eds., *The Handbook of Social Psychology*, vol.2 (2nd ed.), Addison-Wesley, Reading, Massachusset (USA), p. 693-728.
- Zaheer S., (1995), Overcoming the Liability of Foreignness, *Academy of Management Journal*, 38, No.2, p. 341-363.



Annexe 1 : Questionnaire en anglais

Please circle the number indicating your opinion of your firm. Alternative 4 indicates that no changes have occurred.			
1-Growth opportunities in the environment	<i>Have decreased dramatically</i>	1-2-3-4-5-6-7	<i>Have increased dramatically</i>
2-Production/ Service Technology in our principal industry	<i>Has changed very much</i>	1-2-3-4-5-6-7	<i>Has remained the same</i>
3-Rate of innovation of new operating processes in our principal industry	<i>Rate has dramatically increased</i>	1-2-3-4-5-6-7	<i>Rate has fallen dramatically</i>
4-Rate of innovation of new products or services in our principal industry	<i>Rate has dramatically increased</i>	1-2-3-4-5-6-7	<i>Rate has fallen dramatically</i>
5-Research and development activity in our principal industry	<i>Has substantially increased</i>	1-2-3-4-5-6-7	<i>Has fallen off greatly</i>
6-Market activities of our key competitors	<i>Now affect our firm in many more areas (pricing, marketing, delivery, service, production, quality, etc.)</i>	1-2-3-4-5-6-7	<i>Now affect our firm in far fewer areas</i>
7-Market activities of our key competitors	<i>Have become far less predictable</i>	1-2-3-4-5-6-7	<i>Have become far more predictable</i>
8- Market activities of our key competitors	<i>Have become far more hostile</i>	1-2-3-4-5-6-7	<i>Have become far less hostile</i>
9-Are they great differences amongst the Products/ services you offer, with regard to customers' buying habits	<i>About the same for all our products</i>	1-2-3-4-5-6-7	<i>Varies a great deal from one line to another</i>
10-Are they great differences amongst the Products/ services you offer, with regard to the market dynamism and uncertainty	<i>About the same for all our products</i>	1-2-3-4-5-6-7	<i>Varies a great deal from one line to another</i>
11- Legal, political and economic constraints (e.g. government regulations)	<i>Have proliferated greatly</i>	1-2-3-4-5-6-7	<i>Almost no change</i>
12-Are they great differences amongst the Products/ services you offer, with regard to the nature of competition	<i>About the same for all our products</i>	1-2-3-4-5-6-7	<i>Varies a great deal from one line to another</i>



Annexe 2 : Questionnaire final en français

De quelle façon pourriez-vous décrire l'environnement économique de votre entreprise

	Pas du tout <i>d'accord</i>						Tout à fait <i>d'accord</i>
1 - Les possibilités de croissance dans cet environnement ont augmenté considérablement	1	2	3	4	5	6	7
2 - Vous n'avez pas été amené à changer votre technologie de production et de gestion dans votre activité principale	1	2	3	4	5	6	7
3 - Le rythme de l'innovation dans les procédés de production s'est accéléré considérablement	1	2	3	4	5	6	7
4 - Le rythme de l'innovation en nouveaux produits et services s'est accéléré considérablement	1	2	3	4	5	6	7
5 - Les activités de recherche et de développement ont augmenté substantiellement	1	2	3	4	5	6	7
6 - Les démarches commerciales de vos principaux concurrents affectent votre entreprise dans beaucoup de nouveaux domaines (par exemple barèmes de prix, délais et modalités de livraison, services commerciaux, qualité des produits, ...)	1	2	3	4	5	6	7
7 - Les démarches commerciales de vos principaux concurrents ne peuvent plus être anticipées facilement	1	2	3	4	5	6	7
8 - Les démarches commerciales de vos principaux concurrents sont devenues plus agressives	1	2	3	4	5	6	7
9 - Une plus grande diversité des produits et des services que vous offrez est nécessaire afin de les adapter à vos différents clients	1	2	3	4	5	6	7
10 - Une plus grande diversité des produits et des services que vous offrez est nécessaire afin de l'adapter au dynamisme du marché et à ses incertitudes.	1	2	3	4	5	6	7
11 - Les pressions juridiques (accords, conventions, lois) et économiques se sont fortement accrues	1	2	3	4	5	6	7
12 - Une plus grande diversité des produits et des services que vous offrez est nécessaire afin de répondre à la concurrence	1	2	3	4	5	6	7