

Comparabilité et équivalence des mesures dans les études cross-  
culturelles : des enjeux pour la recherche en logistique

Jean-Louis Chandon<sup>a</sup>, Boris Bartilowski<sup>b</sup>,

a Professor of Marketing, INSEEC BUSINESS SCHOOL, 27 Avenue de Vellefaux, Paris 75010, phone +33 6 75 89 66 97, email: [jichandon@inseec.com](mailto:jichandon@inseec.com)

b KEDGE Business School, Domaine de Luminy, BP 921, 13288 Marseille Cedex 9, France  
[boris.bartilowski@kedgebs.com](mailto:boris.bartilowski@kedgebs.com)

An ulterior version of this Working Paper will be published later on in “Images de la logistique: éclairages managériaux et sociétaux”, Presses Universitaires d'Aix-Marseille Editions, JL. Chandon, B. Bartilowski, 2017.

## Comparabilité et équivalence des mesures dans les études cross-culturelles : des enjeux pour la recherche en logistique

Jean-Louis Chandon, INSEEC Business School - Boris Bartikowski, Kedge Business School

### Introduction

Souvent les chercheurs sélectionnent un instrument de mesure développé dans un autre pays, comme les Etats-Unis par exemple, et se contentent de le traduire pour l'appliquer dans différents pays cibles.

Cependant, comment adapter un instrument de mesure développé dans un pays donné, à l'intérieur d'une culture donnée afin de le transposer à un autre pays ou à une autre culture ? Comment s'assurer qu'il sera compris de la même manière, que les échelles (items ou objets support de la mesure) seront utilisées de manière identique par les répondants et que les qualités initiales de fiabilité et de validité seront maintenues dans une autre culture ? C'est là tout le problème de l'équivalence cross-culturelle des instruments de mesure. Il est nécessaire d'examiner la comparabilité et l'équivalence de l'échelle transposée avec l'échelle initiale, faute de quoi on s'expose au risque de croire que les différences observées sont factuelles alors qu'elles ne sont que le reflet de l'absence de comparabilité ou d'équivalence entre les versions transposées de l'échelle originale. Bauer (1989) propose une distinction entre comparabilité des mesures et équivalence des mesures qui est reproduite dans la Figure 1. La comparabilité et l'équivalence des mesures doivent être démontrées avant toute analyse cross-culturelle, et les recherches internationales en logistique et SCM doivent tenir compte de ce point essentiel.

**Figure 1.** Equivalence et comparabilité des mesures en recherches culturelles (d'après Bauer, 1989)

<b>Comparabilité des mesures</b>	
Définition du problème	Equivalence fonctionnelle, conceptuelle et catégorielle
Méthodes de collecte	Mêmes procédures de collecte et mêmes stimuli
Unité de recherche	Même définition et sélection des unités de recherche
Administration	Même procédures, durées et interactions
Enregistrement	Mêmes catégories et même codage

<b>Equivalence des mesures</b>	
Invariance de configuration	Les mêmes items mesurent les mêmes concepts
Invariance Métrique	Les poids factoriels ( <i>loadings</i> ) sont équivalents
Invariance scalaire	Mêmes ordonnées à l'origine et même poids factoriels

Les recherches comparatives transculturelles sont parmi les plus anciennes des approches scientifiques en sciences sociales (Durkheim, 1895). La recherche comparative a été reconnue comme la base de la méthode scientifique (Campbell et Stanley, 1963). Les comparaisons culturelles permettent d'observer ou d'inférer les différences, similitudes, variations, covariations et causalité (Berry, 1980). La disponibilité croissante de grandes enquêtes internationales ouvre la voie des comparaisons transculturelles et longitudinales pour les économistes, les gestionnaires et les sociologues. A titre d'exemple, on peut citer les enquêtes telles que European Social Survey (ESS), International Social Survey Program (ISSP), étude de valeurs européennes (SVE), World Values Study (WVS), European Household Panel Study, le programme d'évaluation des performances internationales des élèves et le Global

Entrepreneurship Monitor. On peut aussi citer l'existence de baromètres régionaux tels que Afrobarometer, Asianbarometer, Eurobarometer et Latinobarometer. Ces enquêtes et baromètres mesurent divers concepts théoriques comme les valeurs, les dispositions psychologiques ou divers schémas comportementaux au moyen d'indicateurs multiples dans le but de faire des comparaisons entre différents pays, régions et points dans le temps. Ces comparaisons sont nécessaires pour faire progresser les sciences sociales.

Le domaine de la logistique n'échappera pas à cette tendance. Par exemple, il sera intéressant d'évaluer comment des décideurs logistiques, en fonction de leur culture, perçoivent, voire interprètent, la pression exercée en termes de niveau de service imposée par une mise en tension des flux. Dans certaines zones géographiques, un fournisseur pourra, par exemple, considérer qu'une livraison sous huit jours de 90 % des composants est une performance tout à fait acceptable, tandis que dans une autre zone géographique, une telle performance sera considérée comme médiocre. En termes de valeurs, il pourrait également être important de mesurer des niveaux d'opportunisme dans les relations entre vendeurs et acheteurs de services logistiques : là où, dans une culture donnée, le fait que le vendeur ne dévoile pas totalement ses capacités de transport et de stockage sera considéré comme faisant partie des « règles du jeu » commercial, dans une autre culture, il s'agira du signe fort d'un comportement déviant aboutissant à se méfier du partenaire, et à ne pas s'engager avec lui dans un partenariat durable.

Cependant, obtenir et analyser les données transnationales soulève de réelles difficultés. Les répondants — la source d'information dans les sondages — peuvent ne pas fournir de réponses correctes, peuvent dire une chose, mais en faire une autre, peuvent répondre d'une manière qu'ils estiment socialement souhaitable ou être facilement influencés par les situations contextuelles. En outre, les répondants dans les enquêtes transnationales ont été socialisés dans milieux économiques et culturels très divers, parlent des langues différentes et peuvent comprendre certaines idées ou concepts de diverses manières et le plus souvent d'une manière culturellement spécifique.

De nombreuses recherches dans le domaine du management, des ressources humaines et du marketing développent une large variété de comparaisons entre divers pays ou entre diverses cultures à l'intérieur d'un même pays. De manière parallèle, des travaux en logistique et *supply chain management* (SCM) ont intégré la dimension culturelle (McAfee et al., 2002 ; Mello et Stank, 2005 ; Jia et Rutherford, 2010 ; Davis et al., 2014).

## **1. Comparabilité des mesures**

La première question concerne la *définition des concepts* et du thème de recherche. Les construits manipulés fonctionnent-ils de manière identique dans les deux cultures ? En suivant les travaux de Bauer (1989) et de Malhotra (1996), on distingue, dans la phase de définition du problème de recherche, trois aspects de l'équivalence du construit, à savoir l'équivalence fonctionnelle, l'équivalence conceptuelle et l'équivalence catégorielle :

- On parle d'*équivalence fonctionnelle* quand on est capable de mesurer le même comportement ou la même attitude ou perception dans des sociétés différentes d'un pays ou d'une culture à une autre (Kline, 1983). Par exemple, Chandon, Laurent et Kapferer (2016) proposent que les consommateurs de produits de luxe sont capables d'indiquer leur perception du prix à partir duquel un produit est considéré comme luxueux.

- L'*équivalence conceptuelle* concerne les interprétations portées par les individus sur les objets, les stimuli ou les comportements. On doit considérer que ce qui apparaît comme ayant un sens évident pour une personne donnée d'un pays donné peut dépendre entièrement des normes culturelles. Les objets, les stimuli ou les comportements existent-ils de la même manière et sont-ils exprimés de la même manière dans les cultures ou les pays ? Si oui, l'équivalence conceptuelle est acquise. On peut évaluer l'équivalence conceptuelle en examinant la validité du construit dans les différentes cultures selon les procédures classiques telles que décrites par Babin (1994); Grunert et al. (1993) ou Churchill (1979).
- Enfin, l'*équivalence catégorielle* examine les regroupements entre idées et comportements (Mintu et al., 1994). Par exemple, Chandon, Laurent et Kapferer (2016) montrent que les produits de luxe personnels sont classés dans les mêmes trois familles de produits par les répondants de sept pays.

La deuxième question concerne les *méthodes de collecte des données*. Sont-elles comparables d'un pays à un autre ? Les unités de recherche sont-elles identiques et sélectionnées selon le même processus ? La qualité du processus de collecte et de contrôle des données est-elle comparable ? La manière dont les répondants réagissent au questionnement conduisent-elles à des distorsions similaires ou bien divergent-elles d'une culture à l'autre ? Les échantillons jouissent-ils d'un même degré de représentativité d'un pays à l'autre ? Tous ces aspects nécessitent des contrôles précis. Par exemple, Laurent et Kapferer (2016) et Chandon, Laurent, Kapferer (2016) n'interrogent dans les divers pays que les répondants ayant acheté pour eux même ou pour autrui, ou ayant reçu en cadeau des produits de luxe personnels au cours de l'année passée, ceci afin d'assurer un minimum de familiarité avec les produits de luxe.

En ce qui concerne l'*équivalence au niveau de l'échantillon*, on distingue (a) l'équivalence de la structure de l'échantillon (par exemple en termes de critères sociodémographiques) et (b) l'équivalence de la taille (par exemple afin d'obtenir un niveau de puissance des test comparable d'un pays à un autre). Pour l'équivalence de l'instrument de mesure, on retient l'équivalence de traduction (Brislin, Lonner et Thorndike, 1973, Broderick, 1999 ; Douglas et Craig 1983). Celle-ci peut être subdivisée en (a) équivalence lexicale, (b) équivalence idiomatique et (c) équivalence grammaticale et syntactique. Enfin, l'équivalence au niveau de la procédure de mesure se subdivise en (a) équivalence de la coopération des répondants, (b) équivalence contextuelle, (c) équivalence du style des réponses. La traduction des échelles suivie de la retraduction dans la langue d'origine de l'échelle ne résout pas tous les problèmes, comme le montrent Chandon et Bartikowski (2010).

Il est bien connu que les répondants ne répondent pas toujours en toute vérité mais choisissent la réponse qui les mettra le mieux en valeur car certaines réponses sont plus socialement désirables que d'autres. Ainsi, le représentant commercial d'un prestataire de services logistiques (PSL), face à un client chargeur, garantira qu'il est capable de gérer les stocks de son client avec un taux de rupture équivalent au standard du secteur, voire meilleur. Mais ce qui est socialement désirable varie d'une culture à une autre. Depuis les travaux de Paulhus (2002), on admet l'existence de deux facettes de cette tendance. Une facette appelée « Egoistic Response Tendency » (ERT) par laquelle certains répondants cherchent à faire valoir leur autonomie, leur influence sur les autres, leur statut, leur indépendance (je suis un super héros) et une facette appelée « Moralistic Response Tendency » (MRT) par laquelle d'autres répondants mettent l'accent sur leur affiliation, leur appartenance, leurs connexions, leur amour du prochain (je suis un saint).

Les travaux récents de Steenkamp, De Jong et Baumgartner (2010) portant sur 12 424 répondants répartis dans 26 pays et 4 continents montrent que le Royaume-Uni et l'Irlande se situent au niveau le plus bas sur ces deux facettes de la désirabilité sociale tandis que les Chinois sont les champions du MRT et les Hongrois les champions de l'ERT. Les Etats-Unis se situent légèrement au-dessus de la France pour les deux facettes, tandis que l'Allemagne et la Suisse sont proches du Royaume Uni. L'Espagne et l'Italie ont des scores ERT élevés. La Russie, la Thaïlande, l'Argentine et la Pologne ont des scores MRT élevés. Les travaux de Crowne et Marlowe (1964) montrent qu'il existe une forte association entre MRT et l'échelle individualisme / collectivisme de Hofstede. Le score MRT est plus élevé dans les pays de culture collective que dans les pays de culture plus individualiste. Il convient aussi de prendre en compte les styles de réponse, c'est-à-dire les tendances à répondre indépendamment du contenu de la question posée. Le tableau 1 adapté de Baumgartner et Steenkamp (2001) décrit six styles de réponse.

**Tableau 1.** Six styles de réponse

Style de réponse	Définition
Approbation ( <i>yea-saying</i> )	Tendance à approuver tous les énoncés
Désapprobation ( <i>nay-saying</i> )	Tendance à désapprouver tous les énoncés
Extrémisme	Tendance à répondre aux extrêmes de l'échelle
Centrisme	N'utiliser qu'une faible partie (centrale) de l'échelle
Désengagement	Tendance à n'utiliser que le point milieu de l'échelle
Réponse aléatoire	Absence de motivation ou incompréhension

Les auteurs montrent que parmi 11 pays européens, et pour 14 échelles de mesure, les styles de réponse affectent diversement les échelles et les pays. Les deux pays présentant la plus forte contamination des réponses sont la Grèce et le Portugal. Au-delà des styles de réponse, les répondants peuvent choisir une catégorie de réponse spécifique car elle présente une importance particulière dans leur culture. Par exemple, le chiffre « 3 » est associé à la chance pour les Chinois, alors que le chiffre « 4 » porte malheur. Ces préférences numériques peuvent biaiser les comparaisons interculturelles. En logistique, le « respect des délais » ne renverra ainsi pas toujours à une signification équivalente que l'on soit dans un pays latin ou dans un pays nordique. Les quelques exemples ci-dessus montrent qu'avant d'aborder l'équivalence statistique des mesures il convient de contrôler strictement leur comparabilité, condition *sine qua non* de leur possible équivalence.

## 2. Equivalence des mesures

Plusieurs méthodes quantitatives permettent d'évaluer l'équivalence des mesures. La première famille de méthode repose sur l'analyse factorielle confirmatoire multi-groupe fondée par Jöreskog (1971) et Sörbom (1974), puis développée par Horn et McArdle (1992). Elle a été appliquée en marketing dans les travaux pionniers de Mullen (1995), Singh (1995), et Steenkamp et Baumgartner (1998), et en ressources humaines par James et Brett (1994). Elle vient de connaître de nouveaux développements importants avec l'apparition du concept d'invariance partielle initié par Byrne, Shavelson et Muthén (1989), puis développé par Asparouhov et Muthén (2014), et disponible dans le logiciel MPLUS version 7 avec la procédure d'alignement. La seconde famille de méthodes utilise l'analyse des *courbes de réponse* aux items (Item Response Theory) développée par Rasch (1960), Lord et Novick (1968) et Samejima (1969). La troisième famille de méthode est plus ancienne. Elle remonte à

Thurstone (1928) et Thurstone et Chave (1929). Elle repose sur la *calibration culturelle* préalable des échelles avant leur utilisation et nécessite de recourir à deux échantillons, l'échantillon de calibration et l'échantillon de mesure.

**2.1 - L'analyse factorielle confirmatoire multi-groupe** s'applique lorsque la relation entre les énoncés et les concepts latents est supposée linéaire et qu'elle résulte d'analyses factorielle exploratoires (EFA) et confirmatoires (CFA), fondées sur une théorie et une sélection d'énoncés provenant de la culture de départ. La première question qui se pose est la stabilité du modèle de mesure d'une culture à une autre. Pour démontrer cette stabilité ou invariance culturelle, on effectue une série de tests sur des modèles hiérarchiquement emboîtés, dont on compare la qualité d'ajustement au moyen de divers tests (Delta  $X^2$ , Delta CFI, Delta RMSEA ou Delta SRMR) par rapport à un modèle initial non contraint.

L'*invariance de configuration* (Horn et McArdle 1992), requiert que les mêmes énoncés soient en relation avec les mêmes concepts quel que soit la culture ou le pays. C'est un critère structurel. Il permet que les corrélations entre les énoncés et le concept puissent varier légèrement à condition que l'on ne puisse pas détecter de différences d'affectation d'un énoncé à un concept. L'invariance de configuration interdit qu'un énoncé attaché au concept A dans la culture de départ se trouve attaché à un concept B dans la culture de transposition ou encore qu'un énoncé bien relié au concept dans la culture de départ (*loading* supérieur à 0,7) deviennent insatisfaisant dans la culture de transposition (*loading* inférieur à 0,5). De tels énoncés devront être éliminés. Si la contrainte d'invariance de configuration ne réduit pas de manière significative la qualité d'ajustement par rapport au modèle non contraint, l'invariance de configuration est démontrée et devient la base de comparaison pour des modèles plus contraint. On peut alors passer au test de l'invariance métrique. Par contre, si l'invariance de configuration n'est pas vérifiée, la comparaison inter culturelle est impossible et l'échelle de mesure initiale ou le modèle théorique initial doivent être amendés.

L'*invariance métrique* est un critère d'équivalence plus strict. Il requiert que les corrélations entre les énoncés et les concepts qu'ils reflètent soient non significativement différentes d'une culture à une autre. L'invariance métrique peut être démontrée par un test Delta  $X^2$  mais la sensibilité de ce test à la taille des échantillons amène à préférer les tests Delta CFI, Delta RMSEA ou Delta SRMR (Cheung, Rensvold, 2009).

L'*invariance scalaire* ajoute à l'invariance métrique l'égalité, d'une culture à l'autre, des ordonnées à l'origine dans la relation linéaire entre les énoncés et les concepts qu'ils reflètent. En d'autres termes, il s'agit de prouver que les énoncés ont la même « difficulté » d'une culture à l'autre. En pratique, le critère d'invariance scalaire se révèle extrêmement strict et n'est presque jamais atteint. On se contente alors d'une invariance scalaire *partielle* (Steenkamp et Baumgartner, 1998 ; Byrne et al, .1989), en établissant la liste des énoncés qui vérifient cette propriété et la liste de ceux qui ne la vérifient pas. Si au moins deux énoncés vérifient l'invariance scalaire l'analyse peut se poursuivre. Lorsque l'invariance de la mesure n'est pas satisfaite, on doit trouver des sous-groupes de pays qui sont invariants.

Le niveau d'invariance requis dépend de l'objectif de l'étude. L'invariance de configuration peut suffire s'il s'agit d'explorer la structure des construits d'une culture à une autre. Par contre, l'invariance métrique est requise pour comparer des modèles structurels (Byrne et al. 1989 ; Steenkamp et Baumgartner 1998). Si l'on désire comparer les moyennes des concepts d'un pays à l'autre, alors l'invariance scalaire, au minimum partielle, est nécessaire. Pour ces trois niveaux d'invariance, Chen (2007) suggère de rejeter l'invariance si

Delta CFI > 0,01 et Delta RMSEA > 0,015. Dans le cas de conclusion divergente des deux tests, le test Delta CFI doit être choisi comme critère principal.

L'analyse factorielle confirmatoire présente plusieurs limites. Avant la disponibilité de la procédure d'alignement dans le logiciel MPLUS 7, le chercheur souhaitant aboutir à l'invariance scalaire partielle devait examiner manuellement un grand nombre de modifications du modèle de mesure et s'exposait ainsi à la capitalisation sur la chance comme indiqué par MacCallum, Roznowski et Necowitz (1992). De plus, les échelles les plus fréquemment utilisées sont des échelles de Likert avec 5 à 7 degrés d'accord. En toute rigueur, ces échelles présentent un niveau de mesure ordinal. Il convient donc de les utiliser non pas avec les estimations classiques par maximum de vraisemblance mais avec des techniques d'estimations appropriées, fondées sur les corrélations polychoriques, ce qui complique les comparaisons multi groupe, comme le précisent Lubke et Muthén (2004).

Une revue de la littérature interculturelle menée par He, Merz et Alden (2008) portant sur 243 articles publiés dans 15 revues de premier rang de 2000 à 2005 indique que seulement 28% de ces articles présentent un test d'invariance des échelles de mesure. Plusieurs raisons peuvent expliquer cette carence. Tout d'abord, les analyses d'invariance sont d'une redoutable lourdeur lorsque les données portent sur un grand nombre de pays. Ensuite, les chercheurs ne disposent pas toujours d'échantillons de taille suffisante dans chacun des pays (Bagozzi et Yi, 1989 ; Bollen 1989 ; Myers et al., 2000). Enfin, les études utilisant un seul item pour mesurer un concept ne peuvent pas faire l'objet de test d'invariance.

**2.2 - L'analyse des courbes de réponse aux énoncés** (Item Response Theory ou IRT) offre une approche complémentaire. L'analyse IRT étudie la courbe de réponse de chaque énoncé au concept latent à mesurer mais, à la différence des méthodes factorielles classique, la réponse est non linéaire et probabiliste. Développée initialement pour des énoncés de nature binaire visant à mesurer une aptitude ou une compétence supposée uni dimensionnelle, la méthode a été étendue par Samejima (1969), aux énoncés comportant à plusieurs niveaux d'accord (échelles ordinale ou échelle d'intervalle) et aux concepts comportant plusieurs dimensions. De nouveaux développements (DIF trees) étendent l'analyse aux comparaisons interculturelles. Les tests DIF (Differential Item Functioning) examinent, énoncé par énoncé, si les répondants ont des probabilités de succès différentes selon leur pays d'origine (Borsboom, Mellenbergh et van der Linden, 2002). Comme pour les analyses factorielles l'analyse des courbes de réponse permet de détecter les énoncés invariants d'une culture à une autre. Si l'objectif est de procéder à des comparaisons moyennes, le modèle IRT peut être utilisé immédiatement car les démonstrations d'invariance de mesure ne sont plus nécessaires puisque les scores sur le concept latent ont tous à la même échelle (De Jong, Steenkamp, Veldkamp, Fox, 2007). Si l'objectif est de comparer les relations structurelles, les chercheurs peuvent utiliser une approche en deux étapes, en estimant d'abord les scores factoriels puis en les introduisant ensuite dans les techniques de régression / ANOVA pour estimer leurs hypothèses de fond.

### **2.3 – La calibration des échelles sémantiques par la loi des « jugements catégoriels.**

Cette méthode est plus rarement utilisée et on peut le regretter car c'est souvent le choix « des mots pour le dire » qui permet de capter les nuances fines d'un concept. Bartikowski, Chandon, Gierl (2006) résument la méthode en six étapes :

1. Identification d'échantillons d'énoncés dans chacun des pays étudiés.
2. Traduction non littérale des énoncés en recherchant les énoncés qui bien que lexicalement différents produisent des intervalles égaux dans la culture du pays.

3. Calibration des énoncés par des juges dans chacun des pays
4. Suppression des énoncés à forte variance selon l'évaluation des juges.
5. Estimation de la largeur catégorielle des énoncés retenus dans chacun des pays.
6. Demander aux répondants de choisir parmi les énoncés calibrés ceux qui reflètent le mieux leur opinion ou leur attitude.

## Conclusion

Globalement, il apparaît que beaucoup de comparaisons culturelles dans le domaine du marketing et de la logistique gagneraient à mettre en œuvre soit des analyses factorielles multi-groupes afin d'évaluer l'invariance de leurs modèles de mesure soit d'utiliser l'analyse des courbes de réponse aux énoncés. Enfin si l'on envisage l'utilisation d'échelles sémantiques, leur calibration internationale est une précaution utile et nécessaire. Il s'agit là de pistes fécondes à explorer dans les travaux en SCM pour sortir d'un syndrome universaliste qui tend à imposer une vision standardisée du fonctionnement des chaînes logistiques, indépendamment des cultures dans lesquelles elles se meuvent.

## Bibliographie

- Asparouhov, T., Muthén B. (2014), Multiple-Group Factor Analysis Alignment, Structural Equation Modeling
- Babin, L. (1994), Guidelines for State-of-the-Art Scale Development, Proceedings of the Southern Marketing Association Annual Conference: New Orleans, pp.198-201
- Bartikowski, B., K. Kamei, Chandon J.L. (2010), "A verbal rating scale to measure Japanese consumers' perceptions of product quality." *Asia Pacific Journal of Marketing and Logistics* 22(2): 179-195.
- Vaughn B. K., Wang Q. (2010), DIF Trees: Using Classification Trees to Detect Differential Item Functioning, *Educational and Psychological Measurement*, 70(6) 941–952.
- Bauer, Erich (1989), "Übersetzungsprobleme und Übersetzungsmethoden bei einer multinationalen Marketingforschung," *GfK Jahrbuch der Absatz- und Verbrauchsforschung*, 2, 174-205.
- Baumgartner H., Steenkamp, J.B. (2001), Response Styles in Marketing Research: A Cross-National Investigation. *Journal of Marketing Research (JMR)*, May2001, Vol. 38 Issue 2, p143-156
- Berry, J.W. (1979), Unobtrusive Measures in Cross-Cultural Research, in L. Sechrest (ed.). *Unobtrusive Measurement Today: New Directions for Methodology of Behavioural Sciences*. San Francisco: Jossey-Bass.
- Borsboom, D., Mellenbergh, G. J., & van der Linden, W. J. (2002), Different kinds of DIF: A distinction between absolute and relative forms of measurement invariance and bias. *Applied Psychological Measurement*, 26, 433-450.
- Brislin R.W., Lonner W., Thorndike R.P.M. (1973), *Cross-cultural research methods*, New York, John Wiley & Sons.
- Byrne B., Shavelson R.J., Muthén B. (1989), Testing for the equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance, *Psychological Bulletin*, 105, 456-466.
- Campbell D.T., Stanley J.C. (1963), Experimental and Quasi-experimental Designs for Research, in *Handbook of Research on Teaching*, HOUGHTON MIFFLIN
- Chandon J. L. et Bartikowski B. (2004), Une échelle ordinale permettant de classer les répondants en « satisfait », « indifférent » et « insatisfait », *Recherche et Applications en Marketing*, 19, 1, 39-53.



- Bartikowski B. Chandon J. L. et Gierl H. (2006), Calibration internationale des échelles sémantiques, *Décision Marketing*, 43, 1-13
- Chandon J.L., Bartikowski B. (2011). Les risques liés à la transposition culturelle d'un questionnaire Humanisme et Entreprise, 5-16.
- Chen FF., (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling* 14, 464–504
- Cheung G.W., Rensvold R.B. (2009), Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance, *Structural Equation Modelling*, 233-255.
- Churchill, G.A. (1979), A Paradigm for Developing Better Measures of Marketing Constructs, *Journal of Marketing Research*, XVI, February, pp.64-73.
- Cronbach, L., Meehl, P. (1955): Construct validity in psychological tests. *Psychological Bulletin*, 52: 281-302.
- 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.
- M. Davis, J., Mora-Monge, C., Quesada, G., & Gonzalez, M. (2014). Cross-cultural influences on e-value creation in supply chains. *Supply Chain Management: An International Journal*, 19(2), 187-199.
- De Jong M.G., Steenkamp J.B., Veldkamp B.P. (2009), A Model for the Construction of Country-Specific Yet Internationally Comparable Short-Form Marketing Scales, *Marketing Science*, July 2009, Vol. 28 Issue: Number 4 p674-689
- De Jong M.G., Steenkamp J.B. (2010) Finite Mixture Multilevel Multidimensional Ordinal IRT Models for Large Scale Cross-Cultural Research, *Psychometrika*, March 2010, Vol. 75 Issue: 1 p3-32
- De Jong M.G., Steenkamp J.B., Veldkamp B.P., Fox JP, Baumgartner H. (2007), Using Item Response Theory to Measure Extreme Response Style in Marketing Research: A Global Investigation, *Journal of Marketing Research*, February 2008, Vol. 45 Issue: Number 1 p104-115
- De Jong M.G., Steenkamp J.B., Veldkamp B.P., Fox JP (2007), Relaxing Measurement Invariance in Cross-National Consumer Research Using a Hierarchical IRT Model, *Journal of Consumer Research*, 2007, Vol. 34, p260-278
- De Vecchi D. M. (2002), *Vous avez dit jargon*. Paris: Éditions Eyrolles.
- Douglas S.P., Craig C.S. (1983), *International Marketing Research*, Englewood Cliffs, N-J, Prentice-Hall.
- Durkheim, E. (1938). *The rules of the sociological method*. New York, NY: Free Press [1895].
- Dyer J.H., Chu W. (2000), The determinants of trust in supplier-automaker relationships in the U.S., Japan and Korea, *Journal of International Business Studies*, 31, 2, 259-286.
- He Y., Merz M.A., Alden D.L., (2008), Diffusion of Measurement Invariance Assessment in Cross-National Empirical Marketing: Research Perspectives from the literature and a Survey of Researchers *Journal of International Marketing*, 16, 64-83.
- Horn J.L., McArdle J.J. (1992), A practical and theoretical guide to measurement invariance in aging research, *Experimental Aging Research*, 18, 117-144
- Grunert K.G., Brunsø K., Bisp S. (1993), Food-related life style: Development of a cross-culturally valid instrument for market surveillance, MAPP working paper no 12, October 1993
- Jackson P., Wall T., Martin R. et Davids K. (1993), New measures of job control, cognitive demand and production responsibility, *Journal of Applied Psychology*, 78, 753-762
- James L.R. et Brett J.M. (1984), Mediators, Moderators and tests for mediation, *Journal of Applied Psychology*, 69, 307-321

- Jia, F., & Rutherford, C. (2010). Mitigation of supply chain relational risk caused by cultural differences between China and the West. *The International Journal of Logistics Management*, 21(2), 251-270.
- Kline P. (1983), The Cross-Cultural Use of Personality Tests, In Irvine, S.H. and J.W. Berry (eds.) *Human Assessment and Cultural Factors*: 337-352. New York: Plenum.
- Kapferer J.N., Laurent G. (2015), Where Do Consumers Think Luxury Begin: A Study of Perceived Minimum Price for 21 Luxury Goods in 7 countries, *Journal of Business Research*, 332–340.
- Chandon J.L., Kapferer J.N., Laurent G. (2016), Is the Price Frontier of Luxury Invariant across Countries -- Configural, Metric but not Scalar Invariance, *Monaco Symposium on Luxury*.
- Lubke G.H., Muthén B. (2004), Applying Multigroup Confirmatory Factor Models for Continuous Outcomes to Likert Scale Data Complicates Meaningful Group Comparisons, *Structural Equation Modeling*, 11, 514-534.
- McAfee, R. B., Glassman, M., & Honeycutt, E. D. (2002). The effects of culture and human resource management policies on supply chain management strategy. *Journal of Business logistics*, 23(1), 1-18.
- Malhotra N.K., J. Agarwal and M. Peterson (1996), “Methodological Issues in Cross-Cultural Marketing Research - A State-of-the-Art view”, *International Marketing Review*, 13(5): 7-43.
- Mello, J. E., & Stank, T. P. (2005). Linking firm culture and orientation to supply chain success. *International Journal of Physical Distribution & Logistics Management*, 35(8), 542-554.
- Meredith, W. (1993). Measurement invariance, factor analysis and factor invariance. *Psychometrika*, 58, 525-544.
- Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care*, 44, S69-S77.
- Myers, J. H. et Warner, W. G. (1968), Semantic properties of selected evaluation adjectives. *Journal of Marketing Research*, 5: 409-412.
- Samejima F. (1969), Estimation of latent ability using a response pattern of graded scores, *Psychometrika Monograph Supplement*, 17, 1-100.
- Steenkamp J.B., Baumgartner H. (2010), Assessing Measurement Invariance in Cross-National Consumer Research. *Journal of Consumer Research*, Jun98, Vol. 25 Issue 1, p78-90
- Steenkamp JB, Hofstede F., Wedel M., *Journal of Marketing*, Vol. 63 Issue 2, p55-69
- Paulhus D. (2002), Socially Desirable Responding: The Evolution of a Construct, in *The Role of Constructs in Psychology and Educational Measurement*, Braun H.I., Jackson D.N., Wiley D.E. eds. Mahwah, NJ, Laurence Erlbaum Associates, 46-69
- Peng, T.K., M.F. Peterson and Y.P. Shyi (1991) : Quantitative Methods in Cross- National Management Research: Trends and Equivalence Issues, *Journal of Organizational Behaviour* 12, p : 87-107
- Rasch G. (1960), *Probabilistic Models for Some Intelligence and Attainment Tests*, University of Chicago Press
- Reise S.P., Widaman K.F., Pugh R.H. (1993), Confirmatory Factor Analysis and Item Response Theory: Two Approaches for Exploring Measurement Invariance, *Psychological Bulletin*, 114 (3), 552–66.
- Thurstone, L. L. (1927): Three psychophysical laws. *Psychological Review*, 34: 424-432.
- Thurstone L.L. (1928), Attitudes can be measured, *The American Journal of Sociology*, Vol 26, 249-269
- Thurstone L.L. et Chave E.J. (1929), *The measurement of attitudes*, Chicago, University of Chicago Press

Thurstone, L. L. (1954): The measurement of values. *Psychological Review*, 61: 47-58.

Usunier J.-C. (1999), The use of language in investigating conceptual equivalence in cross-cultural research, in *Seventh Cross-Cultural Consumer and Business Studies Research Conference*, M. Scott (ed.), Cancun, Mexico.