SÉLECTION INTERNATIONALE

Comparaison des validités prédictives des mesures d'un même construit des échelles mono-item et des échelles multi-items

Lars Bergkvist

*Professeur visitant en marketing
Yonsei Graduate School of Business, Séoul*

John R. Rossiter

*Marketing Research Innovation Centre, Université de Wollongong, Australie
Rotterdam School of Management, Erasmus University, Pays-Bas*

RÉSUMÉ

Cette recherche compare les validités prédictives des mesures mono-item et multi-items de l'attitude envers la publicité (A_{Pub}) et de l'attitude envers la marque (A_{Marq}), qui font partie des construits les plus mesurés en marketing. Les auteurs évaluent la capacité de A_{Pub} à prédire A_{Marq} dans des tests de quatre publicités presse concernant différents nouveaux produits. Aucune différence n'est trouvée dans les validités prédictives des mesures mono- et multi-items. Les auteurs concluent que, pour les nombreux construits marketing constitués d'un objet concret unique et d'un attribut concret, tels que A_{Pub} et A_{Marq} , des mesures mono-item doivent être utilisées.

Mots clés : Attitude vis-à-vis d'une publicité, attitude vis-à-vis d'une marque, mesures à un seul item comparées à plusieurs items.

Dans son article très influent publié dans *Journal of Marketing Research*, Churchill (1979, p. 66) déclare la phrase suivante : « En somme, les marketers sont beaucoup mieux servis par les mesures multi-items que par les mesures mono-item de leurs construits et ils seraient avisés de prendre le temps de les développer. » En effectuant cette recommanda-

tion, Churchill suivait la tradition dominante issue de la psychométrie concernant les mesures des capacités et des traits (voir par exemple Guilford, 1954 ; et surtout Nunnally, 1978). Dans les 28 années qui ont suivi l'article de Churchill, les chercheurs ont de plus en plus utilisé des échelles multi-items pour mesurer les construits marketing. Plus précisément, ils ont

Cet article a été publié en mai 2007 dans *Journal of Marketing Research*, 44, 2, 175-184, traduit par Dina Rasolofoarison et reproduit avec la permission de *JMR* et de l'American Marketing Association qui en détient le copyright. L'AMA ne pourra être tenue responsable d'éventuelles erreurs survenues lors de la traduction.

Les auteurs remercient l'aide à la recherche de l'Association of Swedish Advertisers, de la Swedish Newspaper Publishers' Association et de l'Advertising Association of Sweden.

Ils peuvent être contactés aux adresses électroniques suivantes : lars.bergkvist@yahoo.se ; jrossite@uow.edu.au

utilisé des échelles multi-items pour mesurer l'*attribut* des construits (c'est-à-dire l'attitude, la qualité, l'appréciation) qu'il faut différencier de l'*objet* du construit (c'est-à-dire une entreprise, une marque, une publicité). Dans sa procédure C-OAR-SE de développement d'échelle, Rossiter (2002) conclut que si un objet peut être conceptualisé de façon concrète et unique, alors il ne requiert pas plusieurs items pour être mesuré, pas plus qu'un attribut pouvant être défini comme concret. Cependant, l'article de Churchill, comme celui de Peter (1979) sur la fiabilité des échelles multi-items également publié dans *JMR*, ont influencé la mesure des construits marketing à tel point qu'il est virtuellement impossible d'avoir un article accepté en marketing si les concepts principaux ne sont pas mesurés par des échelles multi-items (encore une fois, les items multiples représentant l'attribut du construit). L'utilisation de mesures multi-items est également encouragée par la popularité croissante de l'analyse en équations structurelles (LISREL), un ensemble de techniques statistiques pour lesquelles les mesures multi-items sont la norme, quel que soit le type de construit mesuré (voir par exemple, Anderson et Gerbing, 1988 ; Baumgartner et Homburg, 1996).

Il est en revanche peu commun que les praticiens du marketing utilisent, comme les chercheurs, des échelles multi-items pour mesurer les mêmes construits. Les praticiens semblent préférer les mesures mono-item, non pas pour des raisons théoriques comme celles proposées par Rossiter (2002), auxquelles les praticiens sont peu sensibles, mais pour des raisons pratiques de minimisation des coûts et des refus de la part des répondants. Parmi ces construits communs, on peut trouver l'attitude envers la publicité ou A_{Pub} , que les praticiens appellent l'« appréciation » de la publicité ou AP_{Pub} , et l'attitude envers la marque, que les praticiens et beaucoup de chercheurs appellent « *brand attitude* », symbolisée par A_{Marq} ou parfois A_m . Ces construits populaires sont l'objet d'étude de la présente recherche.

Notre recherche utilise le critère le plus décisif du point de vue de la prise de décision, à savoir la validité prédictive (Aaker *et alii*, 2005) et démontre que des mesures mono-item de ces construits sont tout aussi valides que des mesures multi-items. Le fait que des validités prédictives soient égales signifie que les tests théoriques et les résultats empiriques seraient

similaires si des mesures mono-item étaient utilisées à la place des habituelles mesures multi-items de ces construits. Nous allons maintenant passer en revue les arguments pour et contre les mesures multi-items afin de développer nos hypothèses.

ARGUMENTS EN FAVEUR DES MESURES MULTI-ITEMS

Pour quelles raisons le monde académique estime-t-il que les mesures multi-items sont meilleures que les mesures mono-item ? L'un des arguments théoriques les plus populaires chez les chercheurs trouve sa source dans l'article de Churchill (1979), ainsi que dans celui de Peter (1979), également publié dans *JMR*, sur la fiabilité : les mesures multi-items seraient intrinsèquement plus « fiables » car elles permettent des calculs de corrélations entre les items qui, si elles sont positives et génèrent une corrélation moyenne élevée (c'est-à-dire un coefficient *alpha* élevé), indiquent la « cohérence interne » de tous les items et représentent ainsi l'attribut sous-jacent. Cet argument de la « fiabilité » doit être explicité (voir Rossiter, 2002). Tout d'abord, le coefficient *alpha* ne devrait jamais être utilisé sans avoir au préalable étudié l'unidimensionnalité de l'échelle (Cortina, 1993) ; cela peut être vérifié par une analyse factorielle ou, plus sûrement, par le calcul du coefficient *bêta* de Revelle (1979) qui constitue un bon test d'unidimensionnalité. Une fois l'unidimensionnalité vérifiée, le coefficient *alpha* n'est réellement un indicateur de la validité de mesure par l'ensemble des items que pour un certain type d'attributs, les attributs éliciteurs (voir Rossiter, 2002). Les meilleurs exemples en sont les traits de personnalité (ainsi que les états de court terme correspondants) et les capacités. Le coefficient *alpha* n'est pas approprié pour les deux autres types d'attributs, les attributs « concrets » (qui représentent le type d'attributs utilisés dans la présente recherche dans les construits A_{Pub} et A_{Marq}) et les attributs « formés », tels que la classe sociale (qui est un attribut composé d'une somme de notes sur des variables démographiques). Si

l'attribut du construit est concret, le coefficient *alpha* n'est pas un critère approprié de l'évaluation de la mesure car il n'est pas nécessaire d'avoir plusieurs items pour le mesurer. Un argument logique contre la nécessité d'un coefficient *alpha* élevé est donné par Gorsuch et McFarland (1972), qui mettent l'accent sur le fait qu'une mesure non fiable ne peut pas former une relation n'atteignant pas un haut degré de validité prédictive et qu'ainsi, une mesure mono-item qui aurait une validité prédictive équivalente à une mesure multi-items devrait être considérée comme suffisamment fiable pour remplacer cette dernière. Cronbach (1961, p. 128) dit également : « Si la validité prédictive est satisfaisante, une faible fiabilité ne doit pas nous décourager d'utiliser le test », qui signifie ici la mesure prédictive. Sur la base de ces arguments, nous concluons que la fiabilité ne doit pas entrer en ligne de compte si la mesure mono-item obtient une validité prédictive équivalente à la mesure multi-items.

Un second argument théorique pour l'utilisation d'une mesure multi-items est qu'elle capture davantage d'information qu'une mesure mono-item. Cet argument doit être divisé en deux parties. Premièrement, on considère qu'une mesure multi-items capture davantage d'information qu'une mesure mono-item car elle est « plus à même de capturer toutes les facettes du construit étudié (Baumgartner et Homburg, 1996, p. 143). Cependant, la présence de différences facettes dans un attribut ou un objet signifie que ce construit ne peut être considéré comme un attribut concret d'un objet concret singulier dans la terminologie de Rossiter (2002). Ainsi, cet argument ne concerne pas la présente recherche, étant donné que nous étudions deux construits doublement concrets.

Deuxièmement, on considère qu'une mesure multi-items capture davantage d'information qu'une mesure mono-item car elle offre plus de catégories de réponses. Il est nécessaire d'insister sur le fait que ce n'est pas la multiplicité des items qui est importante ici mais plutôt le nombre de catégories disponibles, ou autrement dit, la « longueur » de l'échelle de réponse. Une échelle multi-items fournit *de facto* une échelle de réponse (potentiellement) plus discriminatoire qu'une mesure mono-item. Par exemple, une mesure de A_{pub} composée de trois items (i) sur des échelles de réponse en sept points (r) peut générer 343 (r^i) réponses uniques différentes et 19 totaux de scores différents ($i \times r - [i - 1]$). Ce nombre

relativement élevé de scores permet de « faire des distinctions relativement fines entre les gens » (Churchill, 1979, p. 66) ou, sur une même ligne, de catégoriser les gens dans un grand nombre de groupes (Nunally et Bernstein, 1994, p. 67). Cet argument est valide tant que le répondant est capable de discriminer un grand nombre de catégories d'un même attribut (Viswanathan, Sudman et Johnson, 2004). Il s'ensuit qu'une mesure multi-items de la variable prédictive devrait montrer une plus grande corrélation avec la mesure de la variable dépendante, soit une plus grande validité prédictive. De plus, découlant du même argument, il s'ensuit que la corrélation entre une variable prédictive mono-item et une variable dépendante multi-items devrait être plus grande que si les deux sont des mesures mono-item, et que la corrélation entre une variable prédictive multi-items et une variable dépendante multi-items devrait être la plus grande de toutes. Cette deuxième partie de l'argument conduit à trois hypothèses.

H_1 : La corrélation entre une variable prédictive multi-items et une variable dépendante mono-item est supérieure à la corrélation entre une variable prédictive mono-item et la même variable dépendante.

H_2 : La corrélation entre une variable prédictive mono-item et une variable dépendante multi-items est supérieure à la corrélation entre la même variable prédictive et une variable dépendante mono-item.

H_3 : La corrélation entre une variable prédictive multi-items et une variable dépendante multi-items est supérieure à la corrélation entre deux mesures mono-item.

ARGUMENTS EN FAVEUR DES MESURES MONO-ITEM

La préférence des praticiens pour les mesures mono-item n'est pas fondée sur des raisons théoriques mais sur des raisons pratiques : des mesures mono-item minimisent le nombre de refus de la part

des répondants et réduisent les coûts de collecte et de traitement des données. Le seul argument théorique (vs empirique) en faveur de l'utilisation de mesures mono-item a été proposé par Rossiter (2002) dans sa procédure C-OAR-SE de développement d'échelle. Rossiter avance qu'une mesure mono-item est suffisante si le construit est, dans l'esprit des évaluateurs (par exemple les répondants d'une enquête), tel que (1) l'objet du construit est « singulier et concret », ce qui signifie qu'il s'agit d'un objet pouvant être facilement et uniformément imaginé, et (2) l'attribut du construit est « concret », ce qui signifie également qu'il peut être facilement et uniformément imaginé. Dans les deux cas, « facilement et uniformément imaginé » est un critère pris de la « théorie de l'image » de Wittgenstein (1961). D'après des jugements d'experts basés sur la procédure C-OAR-SE, A_{Pub} (ou AP_{Pub}) et A_{Marq} sont tous les deux de tels construits.

Un argument empirique pour l'utilisation d'un item unique peut être donné pour les mesures dans lesquelles les multiples items représentant l'attribut (dans la partie réponse de l'item) sont synonymes ou ont pour but d'être synonymes (plus précisément, quand ils sont des adjectifs synonymes). Un exemple extrême est la mesure très connue de l'implication personnelle de Zaichkowsky (1985) (qui, comme construit, se réfère à l'implication personnelle envers un objet, tel qu'une catégorie de produit ou une publicité). Cette mesure utilise 20 paires bipolaires d'adjectifs synonymes pour mesurer l'attribut de « l'implication ». Deux autres exemples sont l'attribut « attitude » des construits A_{Pub} et A_{Marq} tels qu'ils sont mesurés dans la recherche académique. Sur la base d'une étude antérieure de Stuart, Shimp et Engle (1987), Allen (2004) utilise huit paires d'adjectifs synonymes pour mesurer A_{Pub} . C'est un nombre d'items exceptionnellement grand, il est plus fréquent dans la recherche d'utiliser trois ou quatre items synonymes pour mesurer A_{Pub} ou A_{Marq} car c'est un nombre suffisant pour atteindre un coefficient *alpha* élevé. L'argument empirique en faveur de l'utilisation d'un item unique dans de telles mesures est apparu parce que Drolet et Morrison (2001) ont trouvé que l'accroissement du nombre d'items synonymes produit un problème fréquent : plus le nombre d'items synonymes que le chercheur cherche à générer est grand, plus il y a de chances d'inclure des items qui

ne sont pas de parfaits synonymes de l'attribut descriptif original. De plus, ceux qui ne sont pas synonymes ont peu de chances d'être détectés. Drolet et Morrison trouvent que, quand le nombre total d'items augmente, les répondants sont plus enclins à répondre de la même façon à un item qui ne serait pas équivalent (non synonyme, et donc sans validité de contenu) qu'aux autres items de l'échelle. Drolet et Morrison incluent « non-familier/familier » comme item non équivalent dans la batterie des items mesurant A_{Pub} et trouvent que la différence moyenne absolue entre les notations des items équivalents et non-équivalents diminue quand le nombre d'items est augmenté de deux items à cinq puis à dix items (l'un d'entre eux étant l'item non équivalent). La différence moyenne entre le premier item et l'item non équivalent diminue d'approximativement 20 % en passant de deux items à cinq items et d'approximativement 38 % en passant de deux items à dix items. Ces résultats montrent que l'addition de nombreux bons items cache la présence de mauvais items. Si les mauvais items sont corrélés positivement aux bons items, le coefficient *alpha* augmente, ce qui dissuade généralement les chercheurs de rechercher les mauvais items. Paradoxalement, les mauvais items pourraient augmenter la validité prédictive d'une mesure multi-items si la variation des scores des nouveaux items est corrélée à la variation des scores de la variable dépendante, ce qui est probable si le mauvais item est aussi une autre variable prédictive de la variable dépendante. De plus, Drolet et Morrison appliquent la technique de prévisions d'experts pour estimer mathématiquement la valeur informationnelle des items additionnels dans une échelle (voir aussi Morrison et Schmittlein, 1991). En utilisant l'hypothèse des erreurs modérément corrélées, ils montrent que les items additionnels apportent peu d'information ; deux items ayant une corrélation des erreurs de 0,60 apportent l'équivalent de 1,25 item indépendant, et une infinité d'items corrélés à 0,60 apportent autant d'information que 1,67 item indépendant. Ils concluent qu'un ou deux bons items peuvent surpasser une échelle composée de multiples items, si ces multiples items ont une corrélation modérée ou forte de leurs erreurs, ce qui est probable s'ils sont présentés ensemble. L'argument de Drolet et Morrison est entièrement mathématique, ils ne testent pas empiriquement la valeur informationnelle additionnelle des items d'un questionnaire. (Dans la

présente recherche, nous étudions empiriquement la valeur de l'information additionnelle en regardant si des items multiples augmentent la validité prédictive. Si des items multiples ajoutent de l'information, une mesure multi-items d'une variable prédictive devrait prédire les scores d'une variable dépendante avec de plus faibles déviations, résultant ainsi en un r et un R^2 plus élevés.) En raison des problèmes d'erreurs systématiques dans les scores obtenus par des mesures multi-items et de leur démonstration mathématique qui montre que des items additionnels derrière le premier item n'améliorent pas la prédiction des résultats, Drolet et Morrison recommandent l'utilisation de mesures mono-item. Il faut cependant préciser que leur recommandation ne s'applique qu'aux construits qui constituent les classifications les plus basiques des objets et attributs, c'est-à-dire les objets et attributs doublement concrets (Rossiter, 2002).

Un autre argument empirique en faveur des mesures mono-item vient de la volonté d'éviter les biais de similarité des méthodes. Les biais de similarité des méthodes apparaissent quand la corrélation entre deux construits ou plus est augmentée parce qu'ils sont mesurés de la même façon (voir par exemple Williams, Cote et Buckley, 1989). Les biais de similarité des méthodes peuvent apparaître dans les nombreux items d'une mesure multi-items, et peuvent ainsi venir accroître artificiellement son coefficient α . Par exemple, la corrélation entre A_{Pub} et A_{Marq} tendrait à augmenter si chaque construit était mesuré avec plusieurs items de même format (par exemple des items « sémantiques-différentiels ») plutôt qu'avec un item unique de même format. Les biais de similarité des méthodes peuvent aussi accroître la corrélation entre deux mesures mono-item si un format identique est utilisé pour les deux. Pour finir, les biais de similarité des méthodes peuvent également augmenter la corrélation entre deux mesures mono-item si des attributs descriptifs similaires (vs différents) sont utilisés (par exemple « bon/mauvais » pour A_{Pub} et « aime/aime pas » pour A_{Marq}). Ainsi, nous développons trois hypothèses concernant les biais de similarité des méthodes.

H_4 : La corrélation entre deux construits est plus grande si ces construits sont mesurés avec des items multiples de formats identiques que s'ils sont chacun mesurés avec un item unique de format identique.

H_5 : La corrélation entre deux construits est plus grande si ces construits sont mesurés avec des mesures mono-item de formats identiques que s'ils sont mesurés avec des mesures mono-item de formats différents.

H_6 : La corrélation entre deux construits est plus grande si ces construits sont mesurés avec des items uniques employant le même attribut descriptif que s'ils sont mesurés avec des items uniques employant des attributs descriptifs différents.

Le Tableau 1 résume les arguments pour et contre les mesures multi-items et la façon de tester ces arguments si elles existent. Il existe deux tests empiriques importants pouvant être conduits (voir argument 3 dans les deux listes du Tableau 1). L'un des tests est fondé sur l'argument de « discriminabilité » des items multiples et est un test de validité prédictive (H_1 , H_2 et H_3). L'autre test concerne les sources potentielles des biais de similarité des méthodes (H_4 , H_5 et H_6).

Déterminer la validité

Comment un chercheur peut-il décider si une mesure mono-item d'un construit donné est aussi valide qu'une mesure multi-items du même construit ? La procédure C-OAR-SE de Rossiter (2002) affirme que c'est totalement une question de validité de contenu des mesures alternatives. Bien que des entretiens non directifs des consommateurs puissent être nécessaires comme *input*, la validité de contenu est au final déterminée par des jugements d'experts, et non par des recherches quantitatives ou des tests statistiques, à l'exception du calcul de la concordance entre les juges. Cependant, le jugement d'experts n'est pas une option dans la présente recherche. En effet, elle étudie des mesures existantes pour lesquelles les jugements de validité de contenu faits *ex post* n'offriraient pas plus qu'une validité faciale qui n'est pas un type valide de validité car il ne révèle pas les items qui ont été considérés ni ceux qui ont été rejetés, et ainsi ne montre pas comment les items valides ont été sélectionnés (Rossiter, 2002).

La méthode psychométrique habituelle de comparaison de validité consiste à examiner comment chaque mesure prédit les mesures des résultats étu-

Tableau 1. – Arguments pour et contre les mesures multi-items et les façons de les tester

Arguments pour les items multiples	Commentaires	Comment les tester
1. Augmente la fiabilité en permettant le calcul du coefficient <i>alpha</i>	S'applique à tous les construits du paradigme de Churchill (1979). S'applique aux attributs « éliciteurs » selon le paradigme de Rossiter (2002), mais pas aux attributs concrets ou formés	Ne peut pas tester si les attributs sont concrets ou formés. Cela doit être décidé par jugements d'experts. Pour les attributs « éliciteurs », le coefficient <i>alpha</i> peut être calculé quand l'unidimensionnalité a été établie.
2. Nécessaire si l'objet est abstrait ou si l'attribut est abstrait	Les paradigmes de Churchill (1979) et de Rossiter (2002) acceptent cela, bien que la terminologie employée soit de Rossiter. Cependant, le paradigme de Churchill prétend que les items multiples sont nécessaires pour tous les construits pour « capter toutes les facettes du construit ». Cela n'est pas accepté par Rossiter (voir l'argument 1 dans les « Arguments contre les items multiples »).	Ne peut pas être testé. Décidé par jugements d'experts.
3. Capable de mieux discriminer (quand cela est souhaitable) les catégories de l'attribut en augmentant le nombre de catégories dans l'échelle de réponse.	Les deux paradigmes acceptant cela, même si Rossiter (2002) dirait qu'un item unique pourrait être rendu aussi discriminant en augmentant le nombre de catégories dans l'échelle de réponse.	Comparer la prédiction quand la variable prédictive et la variable dépendante sont mesurées avec une échelle multi-items (ex : trois items en sept points, fournissant 19 catégories de réponse possibles) et avec une échelle mono-item (ex : un item en sept points, 7 catégories). Si l'argument de plus grande discrimination et de « plus d'information » est correct, la validité prédictive devrait être la plus élevée pour les mesures multi-items des deux variables, plus basse quand l'une des mesures est mono-item, et la plus basse quand les deux mesures sont mono-item.
Arguments contre les items multiples	Commentaires	Comment les tester
1. Les items multiples ne sont pas nécessaires (pas valides) si l'objet est singulier et concret ou si l'attribut est concret.	La présente étude utilise A_{Pub} , $CROY_{Marq}$ et A_{Marq} . Dans le cadre de Rossiter (2002), chacun a un objet singulier et concret (la publicité ou le produit de marque), et les attributs (croyance ou attitude) sont concrets, donc un item unique devrait être suffisant.	Ne peut pas être testé. Décidé par jugements d'experts.

Tableau 1. – Arguments pour et contre les mesures multi-items et les façons de les tester (suite)

Arguments contre les items multiples	Commentaires	Comment les tester
2. Les items additionnels génèrent le risque de capter un attribut d'une autre variable prédictive.	Selon Rossiter (2002), cela est probable si les items sont censés être des synonymes de l'attribut original.	Décomposer les items des échelles multi-items en variables prédictives indépendantes (par étape) : des items additionnels ne doivent pas significativement augmenter le R^2 (ajusté) s'ils capturent le même attribut. (Il faut noter qu'une « non-différence » dans le test de l'argument 3 de la section « Arguments pour les items multiples » serait aussi une preuve qu'aucun autre attribut n'a été capté.)
3. Biais de similarité des méthodes dans la mesure des variables prédictive et dépendante.	Le biais de similarité des méthodes pourrait augmenter artificiellement la corrélation entre une variable prédictive mono-item et une variable dépendante mono-item mais moins qu'entre deux variables indépendante et prédictive multi-items. Avec une variable prédictive mono-item et une variable dépendante mono-item, les biais de similarité des méthodes peuvent survenir avec l'utilisation d'un format identique ou d'adjectifs descriptifs de l'attribut identiques	Comparer la prédiction quand les variables prédictive et dépendante sont mesurées avec des items multiples de même type (ex: sémantiques différentielles) et quand elles le sont avec des items uniques de même type ; comparer la prédiction quand la variable prédictive utilise le même type de mesure que la variable dépendante mono-item et quand les types de mesures diffèrent (ex : échelles bipolaires pour la variable prédictive et sémantique différentielle pour la variable dépendante) ; comparer la prédiction quand les adjectifs des items uniques des variables prédictive et dépendante sont identiques et quand les adjectifs sont parallèles.

diés (appelée « validité du modèle concurrent » quand les deux mesures sont prises dans la même étude, et « validité prédictive » quand la mesure de la variable dépendante se fait plus tard, mais le terme « validité prédictive » est communément utilisé pour désigner les deux situations). D'un côté, Rossiter (2002) élève une objection à la validité prédictive (voir aussi Borsboom, Mellenbergh et Van Heerden 2004) car l'objectif n'est pas de maximiser la prédiction (maximiser la magnitude de r), mais de mesurer la vraie corrélation (la magnitude de la population R) entre la variable prédictive et la variable dépendante. La vraie corrélation R_{xy} sera généralement considérablement plus petite que 1 car la plupart des résultats ont de multiples causes et qu'en sciences sociales, des corrélations supérieures à 0,6 seraient suspectes (Cronbach, 1961). D'un autre côté, si les deux

variables prédictives comparées sont deux (ou plus) mesures d'un même construit (et donc d'un même attribut), cette objection ne semble plus tenir puisque, bien que la corrélation ne soit pas connue, on peut considérer sans risque que plus la corrélation est haute, plus on est près de la vérité.

Dans la présente recherche, nous employons une analyse de corrélation bivariée et une régression multiple pour comparer les capacités des mesures mono-item et multi-items de l'attitude envers la publicité à prévoir les mesures mono-item et multi-items de l'attitude envers la marque. Premièrement, si l'argument de « pouvoir discriminatoire supérieur » des mesures multi-items est correct, les mesures multi-items de la variable indépendante, de la variable dépendante ou des deux devraient atteindre de plus grands coefficients de validité, r , et de plus grandes parts de

variance expliquées dans les régressions, R^2 , que les mesures mono-item. Deuxièmement, si l'argument des biais de similarité des méthodes est correct, les mesures multi-items devraient produire des prédictions artificiellement augmentées. Cela devrait également être le cas des mesures mono-item utilisant le même type d'échelle de réponse ou les mêmes adjectifs dans les mesures de la variable prédictive et de la variable dépendante.

APPROCHE DE RECHERCHE

Approche générale

Les données de cet article proviennent de questionnaires de consommateurs concernant quatre publicités pour quatre produits différents. Nous avons pré-testé les publicités en utilisant les tests traditionnels (tests de la copie). Nous avons fait tourner l'ordre des publicités. Les participants ont répondu à des mesures mono-item et multi-items pour les mêmes variables de test de publicité, de sorte que la comparaison des méthodes de mesures est basée sur un design intra-sujet plutôt qu'inter-sujet.

Les participants

Les participants étaient des étudiants de première et deuxième année de maîtrise de commerce ayant accepté de participer à « une recherche sur le marketing ». Les participants se sont vu offrir un déjeuner gratuit durant le test de copie, un ticket de loterie de la Croix Rouge ainsi qu'une chance de gagner un ticket de cinéma ou des bons d'achat pour la librairie étudiante. En tout, 92 participants ont rempli le test de publicité, mais la taille des cellules pour les quatre publicités diffère car nous avons filtré les participants qui n'étaient pas intéressés par la catégorie de produit.

La procédure

Quelques semaines avant le test de publicité, nous avons mené une étude visant à mesurer l'intérêt des

participants pour les catégories de produits utilisées dans la recherche. Nous avons mesuré l'intérêt en posant des questions sur les intentions d'achat, les achats réels et les usages de chacune des catégories, même si nous avons éliminé la question sur les achats concernant une des catégories, les plans retraite, car il est évident que des étudiants ne sont pas acheteurs de plans retraite. Afin d'éviter de renseigner les participants sur les catégories de produits utilisées dans la recherche, nous avons posé les mêmes questions sur quatre catégories de produits additionnelles, non utilisées dans la suite de l'étude. Nous avons considéré qu'un participant était intéressé par la catégorie de produit s'il avait répondu positivement à au moins une des trois questions sur l'intérêt (avoir l'intention d'acheter, avoir acheté ou avoir utilisé). L'analyse ne prend en considération que ces participants ayant répondu positivement à au moins une des trois questions. La proportion des participants s'intéressant à une catégorie de produit s'étale de 63 % à 95 %, comme on peut le voir dans les tailles d'échantillon n indiquées dans les tableaux de résultats.

Nous avons mené les tests de publicité par groupes de 25 étudiants (avec des livrets individuels) dans une salle de classe, durant l'heure de déjeuner des étudiants. Chaque participant avait été préalablement assigné à un groupe et avait reçu les instructions pendant le déjeuner. En arrivant, il a été demandé aux participants de s'asseoir, d'attendre et de ne pas regarder le livret posé devant eux. Puis le test a commencé, il a été dit aux participants qu'ils allaient voir quatre publicités pour des marques qui n'étaient pas disponibles sur le marché local mais qu'elles le seraient dans un futur proche. Il leur a également été dit qu'il n'y avait pas de bonne ou mauvaise réponse aux questions posées après chaque publicité et que c'était leur opinion en tant que consommateur et non en tant qu'étudiant en école de commerce qui importait. On a également insisté sur l'importance de répondre à toutes les questions du livret.

Chaque publicité du livret était suivie par toutes les questions lui correspondant. Nous avons fait tourner l'ordre des publicités pour minimiser l'effet de contamination (une analyse de la variance effectuée plus tard a démontré que l'ordre des publicités n'était pas une variable significative). Les participants pouvaient prendre tout le temps qu'ils voulaient pour regarder

les publicités. Les pré-tests ont indiqué qu'un délai de trois minutes était suffisant pour tout le monde.

Matériel

Les publicités étaient des publicités réelles concernant des produits réels, mais aucune des publicités ni aucune des marques concernées n'était disponible sur le marché local dans lequel l'étude a été conduite. Ainsi, les publicités et les marques étaient nouvelles pour tous les participants. Les publicités étaient présentées en couleurs sur une feuille A4 et la qualité du papier et de l'impression était celle d'un magazine. Les marques sur les publicités provenaient de quatre catégories de produits différentes : antidouleur, café, plan de retraite et jeans. Nous avons choisi les quatre catégories de produits *a priori* pour représenter les quatre quadrants de la grille de Rossiter et Percy (1997) : elles étaient respectivement faible implication/informationnelle, faible implication/transformationnelle, forte implication/informationnelle et forte implication/transformationnelle.

Mesures

Le questionnaire contenait les mêmes questions pour toutes les publicités de l'étude. Pour chaque publicité, les participants devaient noter leur appréciation de la publicité (AP_{Pub}), leur attitude envers la publicité (A_{Pub}), leur intention d'acheter la marque (IA_{Marq}), leur attitude envers la marque (A_{Marq}) et leur croyance sur le bénéfice apporté par la marque ($CROY_{Marq}$), dans cet ordre. Les mesures concernant la publicité venaient en premier, immédiatement après l'exposition aux publicités, puis venaient les mesures concernant la marque, posées dans un ordre d'inversion de la « hiérarchie des effets » afin d'éviter que les attitudes et les intentions d'achat ne soient influencées par les croyances (Rossiter et Percy, 1997). Le questionnaire contenait d'autres mesures, telles que les réponses cognitives qui, avec IA_{Marq} , n'ont pas été utilisées dans l'analyse.

Les échelles exactes utilisées pour mesurer les construits dans l'analyse apparaissent dans le Tableau 2. Nous avons tiré les mesures mono-item des mesures multi-items. L'appréciation de la publicité AP_{Pub1} , où

l'indice 1 indique le nombre d'items, était déjà une mesure mono-item utilisée par la plupart des praticiens (Haley et Baldinger, 1991 ; Walker et Dubitsky, 1994). L'attitude envers la publicité, A_{Pub3} , était le troisième item de la mesure utilisée par MacKenzie et Lutz (1989) et a été repris par beaucoup de recherches académiques. Pour la mesure mono-item de l'attitude envers la publicité A_{Pub1} , nous avons sélectionné le premier item bon/mauvais, labellisé $A_{Pub1(G)}$ dans les résultats. L'attitude envers la marque $A_{Marque3}$ était le troisième item de la mesure utilisée par Gardner (1985) et par bien d'autres recherches académiques après lui. Pour étudier les biais de similarité des méthodes avec des mesures mono-item, nous avons sélectionné le troisième item « aime pas/aime » pour faire office de mesure mono-item « différente » de A_{Marq1} , labellisée $A_{Marq1(L)}$; pour la « même » mesure mono-item de $A_{Marque1}$, nous avons sélectionné le même item que pour la mesure mono-item de A_{Pub1} , « mauvais/bon », labellisé $A_{Marq1(G)}$.

En plus de ces construits principaux, les croyances sur les attributs importants de chaque produit, $CROY_{Marq}$, étaient nécessaires pour la régression. $CROY_{Marq}$ était composé des deux à quatre (selon les catégories de produits) attributs les plus importants déterminés par les pré-tests ; nous avons mesuré la force des croyances sur des échelles unipolaires en sept points, allant de « dans une faible mesure » (1) à « dans une très large mesure » (7). Pour chaque produit, nous avons combiné les scores de croyance de chaque attribut dans un index. Nous avons également examiné les régressions produit par produit avec les croyances utilisées comme variables indépendantes séparées, et elles ont donné des valeurs de R^2 presque identiques. Ainsi, nous reportons les résultats calculés sur les index pour gagner de l'espace.

Conformément à la procédure psychométrique classique (voir Cortina, 1993), nous avons analysé les mesures multi-items avec une analyse en composantes principales pour déterminer le nombre de dimensions avant de calculer les coefficients *alpha* ; nous avons trouvé que A_{Pub3} et A_{Marq3} étaient toutes les deux uni-dimensionnelles. Les coefficients *alpha* de ces mesures étaient tous bons ou très bons selon les standards psychométriques acceptés ; ils allaient de 0,85 à 0,93 (voir par exemple DeVellis, 1991 ; Nunnally et Bernstein, 1994).

Tableau 2. – Mesures des principaux construits

Construit	Question	Échelle de réponse
AP_{Pub}	« Concernant la publicité pour /MARQUE/, laquelle des phrases suivantes décrit le mieux vos sentiments envers cette publicité ? »	1. Je l'aime beaucoup 2. Je l'aime bien 3. Ni l'un ni l'autre 4. Je ne l'aime pas beaucoup 5. Je ne l'aime pas du tout
A_{Pub}	« Trois paires d'adjectifs vous sont présentées ci-dessous. Indiquez à quel point l'un ou l'autre des deux adjectifs de chaque paire décrit la façon dont vous percevez la publicité pour /MARQUE/. »	Bonne _ _ _ _ Mauvaise Déplaisante _ _ _ _ Plaisante Défavorable _ _ _ _ Favorable
A_{Marq}	« Trois paires d'adjectifs vous sont présentées ci-dessous. Indiquez à quel point l'un ou l'autre des deux adjectifs de chaque paire décrit votre sentiment général sur /MARQUE/ CATÉGORIE DE PRODUIT/. »	Mauvais _ _ _ _ Bon Plaisant _ _ _ _ Déplaisant Aime pas _ _ _ _ Aime

Notes : Nous avons utilisé des échelles inversées pour la mesure mono-item de AP_{Ad} (ex : 5 = réponse « positive »). Nous avons codé les items multiples de 1 à 7 pour les mesures sémantiques différentielles de A_{Pub} et A_{Marq} (7 = réponse « positive »). Pour les mesures mono-item de A_{Pub} et A_{Marq} , une des trois paires d'adjectifs a été sélectionnée (voir la section « Mesures » dans le texte).

ANALYSE

Approche générale

L'objectif général de l'analyse est de comparer les validités prédictives des mesures multi-items et des mesures mono-item de A_{Pub} et A_{Marq} . La validité prédictive est estimée par deux méthodes. Une méthode compare la corrélation simple bivariee r , entre la variable prédictive (A_{Pub}) et la variable dépendante (A_{Marq}) ; r est la statistique habituelle pour désigner le « coefficient de validité » dans les tests psychométriques pour remplacer, si le critère est mesuré plus tard, la validité prédictive (voir Cronbach, 1961). L'autre méthode est la régression multivariée, qui compare les R^2 . La régression multivariée doit aussi être

étudiée. Fishbein et Middlestadt (1995), parmi d'autres, prétendent que le coefficient de validité (corrélation) entre la variable prédictive et la variable dépendante sera artificiellement augmenté si le modèle causal de la variable dépendante est sous-spécifié, car la variable prédictive unique pourrait inclure les effets d'autres variables causales. Les causes les plus probables pouvant influencer sur A_{Marq} sont censées être A_{Pub} et $CROY_{Marq}$ indépendamment (voir Figure 1), mais si le véritable modèle causal inclut un chemin causal indirect $A_{Pub} \rightarrow CROY_{Marq} \rightarrow A_{Marq}$, le coefficient de régression de A_{Marq} dans le modèle réduit $A_{Pub} \rightarrow A_{Marq}$ sera artificiellement augmenté car il contient en le cachant une part de l'effet de $CROY_{Marq}$ (la part médiatrice) sur A_{Marq} . Cela signifie que si l'effet de $CROY_{Marq}$ est mesuré et partiellement isolé, l'effet de A_{Pub} sera plus faible. En ce qui concerne nos analyses, le modèle théorique exact sous-jacent est sans consé-

 Figure 1. – Modèle causal présumé pour A_{Marq}

quences tant que $CROY_{Marq}$ est inclus en tant que variable prédictive, car la solution statistique à ce problème de variables médiatrices est d'inclure les variables médiatrices et les variables indépendantes dans le modèle de régression (Baron et Kenny, 1986).

Nous avons procédé aux analyses de corrélation et aux régressions séparément pour les quatre publicités de l'étude, car les résultats agrégés auraient été difficiles à interpréter et auraient masqué les différences entre les publicités (ou produits). Nous avons étudié la multicollinéarité dans tous les modèles de régression. Aucun des modèles n'avait d'index de conditionnement plus grand que 15 combiné avec deux ou plus proportions de variance supérieures à 0,90 (Hair *et alii*, 1998), ce qui indique que la multicollinéarité n'était un problème dans aucun des modèles. Dans cette analyse, nous avons testé les significativités des différences des r et des R^2 en utilisant les tests z , suivant ainsi la transformation des corrélations de Fisher (Cohen et Cohen, 1975 ; Howell, 1992).

Mesures mono-item versus multi-items de A_{Pub} comme variable prédictive de la variable mono-item A_{Marq}

La première analyse compare la mesure multi-items de l'attitude envers la publicité A_{Pub3} . Les deux mesures mono-item, la mesure « réduite » $A_{Pub1(G)}$, et la mesure liée AP_{Pub1} , sont des variables prédic-

tives alternatives de la mesure mono-item de la variable dépendante $A_{Marq1(L)}$. Le Tableau 3 affiche les coefficients de validité bivariés, r , et les statistiques multi-variées de validité qui représentent la part de la variance expliquée, R^2 , avec $CROY_{Marq}$ dans les équations de régression.

Sur la base des coefficients de validité bivariés, r , nous pouvons rejeter H_1 . Les mesures mono-item de l'attitude envers la publicité, $A_{Pub1(G)}$ et AT_{Pub1} , étaient toutes les deux équivalentes dans leur prédiction de l'attitude envers la marque, $A_{Marq1(L)}$, et étaient aussi bonnes que la variable prédictive multi-items, A_{Pub3} (pour chaque publicité, aucun r n'était significativement différent des autres, $p > 0,10$).

Les coefficients de validité multivariés, R^2 , qui apparaissent également dans le Tableau 3, révèlent des profils de résultats identiques, rejetant ainsi H_1 . Confirmant nos doutes sur les causes omises, la validité prédictive estimée de A_{Pub} , telle qu'elle était estimée par le coefficient standardisé de régression, était artificiellement augmentée pour trois des quatre produits, à l'exception des plans retraite, quand nous avons omis la variable $CROY_{Marq}$ dans le modèle de régression (ces analyses sont disponibles sur demande). Ainsi, le Tableau 3 fournit les R^2 avec cette variable incluse. La conclusion importante est que le rôle causal de A_{Pub} n'est pas affecté par le fait qu'il soit mesuré par des mesures mono-item ou des mesures multi-items.

Tableau 3. – Coefficients de validité (r) et proportion de variance expliquée dans la régression multiple^a (R^2) pour les mesures mono-item et multi-items de A_{Pub} comme prédicteurs de $A_{marq1(L)}$ mono-item

Produit promu								
Prédicteurs of $A_{Marq(L)}$	Anti-douleur		Café		Plan Retraite		Jeans	
	r	R^2	r	R^2	r	R^2	r	R^2
A_{Pub3}	.75	.58	.77	.72	.68	.48	.68	.58
$A_{Pb1(G)}$	.72	.55	.75	.69	.66	.47	.67	.56
AP_{Pub1}	.74	.58	.73	.67	.60	.40	.68	.62
Tailles des échantillons (n)	80		55		59		86	

^a Les équations de régression incluent $CROY_{Marq}$.

Notes : Tous les r sont significatifs à $p < .01$. Tous les modèles de régression sont significatifs à $p < .01$.

Tableau 4. – Coefficients de validité (r) et proportion de variance expliquée dans la régression multiple^a (R^2) pour les mesures mono-item et multi-items de A_{Pub} comme prédicteurs de A_{marq3} multi-items

Produit promu								
Prédicteurs of A_{Marq3}	Anti-douleur		Café		Plan Retraite		Jeans	
	r	R^2	r	R^2	r	R^2	r	R^2
A_{Pub3}	.80	.69	.75	.80	.72	.52	.65	.61
$A_{Pb1(G)}$	.78	.68	.74	.77	.72	.53	.66	.61
AP_{Pub1}	.77	.67	.70	.76	.68	.49	.62	.62
Tailles des échantillons (n)	80		55		59		86	

^a Les équations de régression incluent $CROY_{Marq}$.

Notes : Tous les r sont significatifs à $p < .01$. Tous les modèles de régression sont significatifs à $p < .01$.

Mesures mono-item versus multi-items de A_{pub} comme variable prédictive de la variable multi-items A_{Marq}

Nous avons répété l'analyse précédente en utilisant la mesure multi-items de l'attitude envers la marque, A_{Marq3} , comme variable dépendante (Tableau 4). Pour les résultats concernant les r , la mesure multi-items de l'attitude envers la publicité, A_{Pub3} , n'était pas significativement une meilleure variable prédictive que les mesures mono-item $A_{Pub1(G)}$ et AP_{Pub1} , et les deux mesures mono-item

n'étaient pas significativement différentes entre elles (pour toutes les publicités, $p > 0,10$). Ces résultats étaient des répliques exactes de ceux concernant la mesure mono-item de la variable dépendante, $A_{Pub1(L)}$. Les résultats concernant les R^2 étaient similaires à ceux concernant les r , montrant ainsi que le rôle causal de A_{Pub} n'est pas affecté selon qu'il est mesuré par une mesure mono-item ou par une mesure multi-items. Parce que les coefficients de validité, r , amenaient dans tous les cas à la même conclusion que les parts de variance expliquée, R^2 , nous nous sommes focalisés seulement sur les r pour tester les cinq hypothèses suivantes.

Pouvoir discriminatoire des mesures mono-item et des mesures multi-items

Les comparaisons des corrélations du Tableau 4 avec celles du Tableau 3 réfutent la position de Churchill (1979) qui affirme que les mesures multi-items sont plus valides car elles capturent une plus grande discrimination dans les réponses du fait du plus grand nombre de catégories de réponse disponibles. Cela ne tenait pour aucun des construits de la présente étude, que ce soit l'attitude envers la publicité (A_{Pub}) ou l'attitude envers la marque (A_{Marq}). Si l'hypothèse était vraie et que les consommateurs n'étaient vraiment capables de mieux discriminer les gradations de leurs attitudes qu'avec des échelles mono-item en sept points, la mesure prédictive multi-items, A_{Pub3} , devrait avoir une corrélation plus élevée avec la mesure dépendante multi-items A_{Marq3} . Même la comparaison la plus extrême, $r_{3,3}$ du Tableau 4 *versus* $r_{1,1}$ du Tableau 3, a révélé que ce n'était pas le cas. Les corrélations étaient de 0,80 contre 0,74 pour la publicité sur les anti-douleur, de 0,75 contre 0,73 pour la publicité sur le café, de 0,72 contre 0,60 pour la publicité sur les plans retraite et de 0,65 contre 0,68 pour la publicité sur les jeans. Bien que les corrélations mono-item soient plus faibles pour les deux produits « informationnels », les anti-douleur et les plans retraite, les corrélations multi-items n'étaient pas significativement plus élevées ($p > 0,10$). Ainsi, les hypothèses H_2 et H_3 étaient rejetées.

Les biais de similarité des méthodes

Les comparaisons des corrélations appropriées allaient à l'encontre de toutes les hypothèses concernant les corrélations artificiellement augmentées dues aux biais de similarité des méthodes. Dans les résultats suivants, aucune comparaison n'était significative ($p > 0,10$). Le résultat montrant que la corrélation multi-items $r_{3,3}$ (Tableau 4) entre A_{Pub} et A_{Marq} n'était pas plus grande que la corrélation mono-item $r_{1,1}$ (Tableau 3) rejette le fait que la répétition d'échelles sémantiques différentielles de ces deux mesures, avec au moins deux répétitions (c'est-à-dire trois items), amène à une prédiction artificiellement augmentée (H_4). Ce résultat élimine un problème potentiel des mesures multi-items, au moins pour celles de moins de trois items.

En ce qui concerne les mesures mono-item, comme le montre le Tableau 5, l'utilisation du même format sémantique différentiel pour la variable prédictive ($A_{Pub1(G)}$) et la variable dépendante n'augmente pas artificiellement les prédictions, en comparaison avec des utilisations de formats différents (AT_{Pub1}), mesuré avec des échelles de réponses « labellisées ». Ainsi, les résultats réfutent H_5 . Enfin, encore pour les mesures mono-item, l'utilisation du même adjectif descriptif pour la variable prédictive et la variable dépendante ($A_{Pub1(G)}$ et $A_{Marq1(G)}$ dans le Tableau 5) n'augmente pas artificiellement la prédiction, en comparaison avec l'utilisation d'adjectifs

Tableau 5. – Coefficients de validité (r) et proportion de variance expliquée dans la régression multiple^a (R^2) pour les mesures mono-item et multi-items de A_{Pub} comme prédicteurs de $A_{marq1(G)}$ mono-item

Prédicteurs of $A_{Marq1(G)}$	Produit promu							
	Anti-douleur		Café		Plan Retraite		Jeans	
	r	R^2	r	R^2	r	R^2	r	R^2
A_{Pub3}	.70	.62	.70	.76	.63	.40	.53	.46
$A_{Pb1(G)}$	.71	.64	.70	.75	.65	.42	.57	.49
AP_{Pub1}	.68	.62	.65	.73	.56	.32	.55	.51
Tailles des échantillons (n)	80		55		59		86	

^a Les équations de régression incluent $CROY_{Marq}$.

Notes : Tous les r sont significatifs à $p < .01$. Tous les modèles de régression sont significatifs à $p < .01$.

différents ($A_{\text{Publ(G)}}$ et $A_{\text{Marq1(L)}}$ dans le Tableau 3). Ces résultats réfutent H_G .

DISCUSSION

Deux des construits les plus largement employés en publicité et en recherche sur le comportement du consommateur sont l'attitude envers la publicité (A_{Pub}) et l'attitude envers la marque (A_{Marq}). Ces deux construits sont doublement concrets (Rossiter, 2002) et ainsi devraient être mesurés de façon valide par un item unique, même si la pratique dominante dans la recherche est de les mesurer avec de multiples items. Dans la présente étude, pour ces deux construits, la mesure mono-item a atteint une validité prédictive équivalente à la mesure multi-items. Nous avons obtenu ce résultat pour quatre catégories de produits différentes et avec deux méthodes différentes d'évaluation de la validité prédictive, le coefficient de validité bivarié, r , et la statistique multivariée, R^2 , quand les construits étaient inclus dans un modèle causal. Ces résultats ne valident pas l'argument psychométrique classique (voir Churchill, 1979 ; Nunnally et Bernstein, 1994) selon lequel les mesures multi-items sont plus valides que les mesures mono-item pour tous les types de construits. En particulier, quand des mesures multi-items sont utilisées pour mesurer des construits doublement concrets, il ne semble pas qu'elles parviennent à mieux les discriminer en capturant plus d'information, ce qui est généralement la justification de leur utilisation.

Nous n'avons trouvé aucune preuve de l'existence des biais de similarité des méthodes avec les mesures multi-items ou avec les mesures mono-item de la variable indépendante ou de la variable dépendante, qui utilisent le même format (dans le cas présent, les échelles sémantiques différentielles) ou le même attribut descriptif (dans le cas présent, « bon/mauvais » pour A_{Pub} et A_{Marq}). Cependant, les mesures multi-items de cette étude étaient constituées de seulement trois items ; cela ne veut pas dire

que les biais de similarité des méthodes n'auraient pas augmenté artificiellement les prédictions quand la variable prédictive, la variable dépendante, ou les deux avaient été mesurées avec un plus grand nombre d'items. Bien qu'il n'y ait pas eu de preuve de cela dans la présente étude n'utilisant que trois items, la multiplicité des items peut produire un effet d'entraînement générant une inflation artificielle de la prédiction (Feldman et Lynch, 1988). Avec plus que trois items, l'accroissement artificiel de la validité prédictive peut avoir lieu.

Une importante limite de nos résultats réside dans le fait que les deux construits utilisés dans notre étude n'ont ni leur objet ni leur attribut constitué de plusieurs composantes, mais ont plutôt un objet concret singulier (la publicité ou la marque) et un attribut concret (l'attitude). La recommandation de la mesure mono-item de A_{Pub} et A_{Marq} ne peut être généralisée à des construits qui ne seraient pas doublement concrets. La théorie de Rossiter (2002) explique que des items multiples sont nécessaires pour mesurer la validité de construits abstraits. Un construit est « abstrait » si (1) l'objet du construit est composé de deux composants ou plus (par exemple le matérialisme, qui est composé de trois composants, à savoir l'utilisation des possessions pour juger du succès d'une personne, la centralité des possessions dans la vie d'une personne et la croyance que les possessions entraînent le bonheur ; voir Richins, 2004), ou est composé d'un ensemble de sous-objets (par exemple, pour la satisfaction au travail, les différents aspects du travail sont le superviseur, les collègues, les responsabilités, la technologie utilisée et les règlements ; voir Gardner *et alii*, 1998 ; Locke, 1969) ou si (2) l'attribut du construit est formé de deux composants ou plus (par exemple la qualité de service avec ses composants de fiabilité, réactivité, empathie, etc. ; voir Parasuraman, Zeithaml et Berry, 1994) et se reflète dans une série d'activités mentales ou physiques (par exemple le trait de personnalité extraversion qui se reflète dans la prise de risque, le caractère grégaire et les activités énergiques ; voir Eysenck, 1967). Les mesures mono-item de ces deux cas d'objets abstraits et de ces deux cas d'attributs abstraits sont certainement moins valides car la signification de l'objet dans les questions mono-item (par exemple « Quelle importante a pour vous le matérialisme ? » « Êtes-vous satisfait de votre travail ? »), ou encore la significa-

tion de l'attribut dans les questions mono-item (« Comment jugez-vous le service chez McDonald's ? » ou « Êtes-vous extraverti ? ») diffèrent beaucoup selon les répondants. Au contraire, les construits abstraits doivent être divisés en composantes concrètes, chacune pouvant être mesurée par un item unique ; il en est de même pour les attributs abstraits. C'est pourquoi nous n'affirmons pas que des mesures mono-item puissent être utilisées de façon fiable pour mesurer n'importe quel construit.

La présente étude peut être considérée comme une extension de la procédure de développement d'échelle de Churchill (1979). Churchill a introduit une approche systématique de développement d'échelle qui a contribué à bâtir les fondements de la méthodologie de recherche en marketing. Il a insisté sur l'importance des considérations théoriques (spécification de domaine) comme étant la première étape de sa procédure. Notre extension concernant les mesures mono-item est essentiellement basée sur, ou limitée par la théorie. Malheureusement, l'insistance de Churchill a reçu beaucoup moins d'attention que ses recommandations sur l'utilisation de mesures multi-items dans la recherche en marketing. Si les chercheurs en marketing avaient fait plus attention à la théorie des construits marketing, ils auraient probablement été moins pointilleux sur l'utilisation de mesures multi-items.

Publicités et marques sont probablement deux des objets les plus utilisés dans les études marketing, que ce soit par les praticiens ou par les chercheurs, et il n'y a aucune raison pour que nos résultats ne soient pas généralisables aux autres objets des études marketing, tels que les entreprises, les distributeurs, les vendeurs, les prix et les promotions des ventes, à partir du moment où ces objets sont concrets et singuliers. De même, l'attitude est l'attribut le plus mesuré en marketing, et nos résultats devraient se généraliser aux autres attributs concrets, tels que les croyances ou les perceptions, les intentions et la satisfaction. Les tests théoriques et les résultats empiriques devraient être inchangés si les bonnes mesures mono-item se substituaient aux habituelles mesures multi-items dans la mesure de ces construits. C'est pourquoi les revues de marketing devraient accepter les articles utilisant des mesures mono-item pour les construits doublement concrets.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Aaker D.A., Kumar V., Day G.S. et Lawley M. (2005), *Marketing research: a Pacific Rim edition*, Milton, Australie, John Wiley & Sons.
- Allen C.T. (2004), A theory-based approach for improving demand artifact assessment in advertising experiments, *Journal of Advertising*, 33 (été), 63-73.
- Anderson J.C. et Gerbing D.W. (1988), Structural equation modeling in practice: a review and recommended two-step approach, *Psychological Bulletin*, 103 (mai), 411-423.
- Baron R.M. et Kenny D.A. (1986), The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations, *Journal of Personality and Social Psychology*, 51 (décembre), 1173-1182.
- Baumgartner H. et Homburg C. (1996), Applications of structural equation modeling in marketing and consumer research: a review, *International Journal of Research in Marketing*, 13 (avril), 139-161.
- Borsboom D., Mellenbergh G.J. et van Heerden J. (2004), The concept of validity, *Psychological Review*, 111 (octobre), 1061-1071.
- Churchill G.A. (1979), A paradigm for developing better measures of marketing constructs, *Journal of Marketing Research*, 16 (février), 64-73.
- Cohen J. et Cohen P. (1975), *Applied multiple regression/correlation analysis for the behavioral sciences*, Hillsdale, NJ, Lawrence Erlbaum Associates.
- Cortina J.M. (1993), What is coefficient alpha? An examination of theory and applications, *Journal of Applied Psychology*, 78 (février), 98-104.
- Cronbach L.J. (1961), *Essentials of psychological testing*, 2^e éd., New York, Harper & Row.
- DeVellis R.F. (1991), *Scale development*, Newbury Park, CA, Sage Publications.
- Drolet A.L. et Morrison D.G. (2001), Do we really need multiple-item measures in service research?, *Journal of Service Research*, 3 (février), 196-204.
- Eysenck H.J. (1967), *The biological basis of personality*, Springfield, IL, Thomas.
- Feldman J.M. et Lynch J.G. (1988), Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior, *Journal of Applied Psychology*, 73 (août), 421-435.
- Fishbein M. et Middlestadt S.E. (1995), Noncognitive effects on attitude formation and change: fact or artifact?, *Journal of Consumer Psychology*, 4, 2, 181-202.
- Gardner D.G., Cummings L.L., Dunham R.B. et Pierce J.L. (1998), Single-item versus multiple-item measurement scales: an empirical comparison, *Educational and Psychological Measurement*, 58 (décembre), 898-915.
- Gardner M.P. (1985), Does attitude toward the ad affect brand attitude under a brand evaluation set?, *Journal of Marketing Research*, 22 (mai), 192-198.

- Gorsuch R.L. et McFarland S.G. (1972), Single versus multiple-item scales for measuring religious values, *Journal for the Scientific Study of Religion*, 11, 1, 53-64.
- Guilford J.P. (1954), *Psychometric methods*, New York, McGraw-Hill.
- Hair J.F., Anderson R.E., Tatham R.L. et Black W.C. (1998), *Multivariate data analysis*, 5^e éd., Upper Saddle River, NJ, Prentice Hall.
- Haley R.I. et Baldinger A.L. (1991), The ARF copy research validity project, *Journal of Advertising Research*, 31 (avril-mai), 11-32.
- Howell D.C. (1992), *Statistical methods for psychology*, 3^e éd., Belmont, CA, Duxbury Press.
- Locke E.A. (1969), What is job satisfaction?, *Organizational Behavior and Human Performance*, 4 (novembre), 309-336.
- MacKenzie S.B. et Lutz R.J. (1989), An empirical examination of the structural antecedents of attitude toward the ad in an advertising pretesting context, *Journal of Marketing*, 53 (avril), 48-65.
- Morrison D.G. et Schmittlein D.C. (1991), How many forecasters do you really have? Mahalanobis provides the intuition for the surprising clemen and winkler result, *Operations Research*, 39 (mai-juin), 519-523.
- Nunnally J.C. (1978), *Psychometric theory*, 2^e éd., New York: McGraw-Hill.
- Nunnally J.C. et Bernstein I.H. (1994), *Psychometric theory*, 3^e éd., New York, McGraw-Hill.
- Parasuraman A., Zeithaml V. et Berry L.L. (1994), Alternative scales for measuring service quality: a comparative assessment based on psychometric and diagnostic criteria, *Journal of Retailing*, 70 (automne), 201-230.
- Peter P.J. (1979), Reliability: a review of psychometric basics and recent marketing practices, *Journal of Marketing Research*, 16 (février), 6-17.
- Revelle W. (1979), Hierarchical clustering and the internal structure of tests, *Multivariate Behavioral Research*, 14, 1, 57-74.
- Richins M.L. (2004), The material values scale: measurement properties and development of a short form, *Journal of Consumer Research*, 31 (juin), 209-219.
- Rossiter J.R. (2002), The C-OAR-SE procedure for scale development in marketing, *International Journal of Research in Marketing*, 19 (décembre), 305-335.
- Rossiter J.R. et Percy L. (1997), *Advertising communications & promotion management*, 2^e éd., New York: McGraw-Hill.
- Stuart E.W., Shimp T.A. et Engle R.W. (1987), Classical conditioning of consumer attitudes: four experiments in an advertising context, *Journal of Consumer Research*, 14 (décembre), 334-349.
- Viswanathan M., Sudman S. et Johnson M. (2004), Maximum versus meaningful discrimination in scale response: implications for validity of measurement of consumer perceptions about products, *Journal of Business Research*, 57 (février), 108-125.
- Walker D. et Dubitsky T.M. (1994), Why liking matters, *Journal of Advertising Research*, 34 (mai-juin), 9-18.
- Williams L.J., Cote J.A. et Buckley M.R. (1989), Lack of method variance in self-reported affect and perceptions at work: reality or artifact?, *Journal of Applied Psychology*, 74 (juin), 462-468.
- Wittgenstein L. (1961), Entry ca. September 29, 1912, in G.E.M. Anscombe et G.H. von Wright (coord.), *Notebooks 1914-1916*, Londres, Basil Blackwell, 7-8.
- Zaichkowsky J.L. (1985), Measuring the involvement construct, *Journal of Consumer Research*, 12 (décembre), 341-352.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.