

L'effet de halo : un biais potentiel dans les méthodes induites de catégorisation des attributs de la satisfaction

Stéphane GANASSALI

I .R.E.G.E.

Université de Savoie - UFR ATE

4, chemin de Bellevue

74 940 ANNECY-LE-VIEUX

France

Tel : +33 450 09 24 00

sgana@univ-savoie.fr

Summary

This article is an investigation of the halo effects occurring in consumer satisfaction surveys, when a derived-importance is measured, using correlation between global and specific satisfactions. We first review several methods of attribute categorization using either stated or derived importance measurements. Then we describe the halo effect that can occur throughout several items in a survey questionnaire. An empirical study is ran, where the halo effect is produced, due to a change in the global assessment position in the questionnaire. Practical implications are finally discussed to prevent the researchers from the use of an inappropriate method.

L'effet de halo : un biais potentiel dans les méthodes induites de catégorisation des attributs de la satisfaction.

La catégorisation des attributs de la satisfaction : un enjeu majeur

Parmi toutes les enquêtes menées chaque année en France, et sur la base d'un observatoire des pratiques mis en place depuis 5 ans auprès de 6 000 organismes par un éditeur de logiciels d'enquêtes, on peut estimer aujourd'hui à plus de 35%, celles qui portent principalement ou exclusivement sur la mesure de la satisfaction des clients, des usagers ou du personnel. Selon l'organisme professionnel SYNTEC Etudes Marketing et Opinion, les études de satisfaction connaissent une très forte croissance depuis 2001 et on estime à environ 10% leur poids relatif sur le marché total des études en France. Ainsi, depuis plusieurs années, portées notamment par la vague des procédures de certification qualité (de type ISO etc.), les enquêtes de satisfaction sont l'objet d'un engouement croissant de la part des organisations de toutes natures. Des enquêtes ponctuelles, nombreuses sont les entreprises qui franchissent le pas et se dotent désormais d'un baromètre de satisfaction, qui est mis en œuvre selon une fréquence pluri-mensuelle ou annuelle.

La recherche académique en marketing a toujours su sur ce point éclairer les pratiques professionnelles en proposant de très nombreux travaux sur le sujet. Certains d'entre eux ont d'ailleurs produit des outils ou des méthodes qui sont aujourd'hui couramment diffusées dans les organisations. On pense bien entendu au modèle SERVQUAL (Parasuraman et al., 1988) ou à l'analyse « Importance-Performance » initialement présentée par Martilla et James (1977). A partir de ces techniques qui sont devenues populaires et dont la pertinence pratique n'est plus à prouver, les chercheurs ont développé de nouvelles approches visant à perfectionner la mesure de la satisfaction surtout mais de l'importance aussi. La recherche sur la mesure de la satisfaction du consommateur au sens large est aujourd'hui un champ très fourni, on compte même depuis 1988, une revue scientifique dédiée exclusivement à ces thématiques : le « Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior ».

Les enjeux stratégiques de l'enquête (ou du baromètre) de satisfaction ne sont plus à démontrer (Ray, 2004). En plus d'un outil de remontée d'information pour la stratégie marketing et le suivi qualité, ces enquêtes s'intègrent en outre dans la politique de gestion de la relation-client. Elles peuvent être utilisées également comme un levier de motivation interne (voire d'évaluation) dans le cadre d'une politique de la gestion des ressources humaines. Comme on le voit, l'enquête de satisfaction peut se situer souvent à un point central de la stratégie générale d'une organisation.

Une enquête de satisfaction comporte généralement une mesure de la satisfaction globale puis une mesure détaillée d'un ensemble de critères identifiés a priori comme les composantes partielles de la perception du client ou du collaborateur. Fréquemment, on questionne également les individus sur l'importance relative qu'ils accordent à chaque item particulier afin de mettre en perspective la satisfaction sur chaque critère avec son importance. Une ou plusieurs questions ouvertes enregistrent au final les suggestions, les commentaires ou les remarques que les répondants souhaitent exprimer sur un des thèmes de l'enquête ou sur tout autre point qui n'aurait pas été abordé préalablement dans le questionnaire. Les variables d'identité du répondant sont enregistrées enfin.

Dans les enquêtes de satisfaction, il est fréquent de s'intéresser non seulement au niveau de la satisfaction, mais aussi à l'importance relative des différents attributs évalués. Cette question étant fondamentale aux yeux de nombreux praticiens, de nombreux travaux se sont intéressés à décrire des méthodes pour mesurer pertinemment l'importance relative des différents critères étudiés dans le cadre d'une enquête de satisfaction. Ces méthodes sont nombreuses, plus ou moins perfectionnées et se différencient selon plusieurs options théoriques et méthodologiques que nous allons rappeler ci-après (voir Windal et Desmet, 2000 ou Bartikowski et Llosa, 2004 pour des synthèses).

Présentation des méthodes directes ou induites

Les méthodes directes

La première différence majeure dans les approches proposées relève tout simplement de la méthode de mesure de l'importance (Chu, 2002). Certains travaux suggèrent une mesure

directe de l'importance sur une base déclarative. C'est donc le répondant qui se prononce sur le poids qu'il accorde à tel ou tel critère dans son évaluation générale de la transaction qu'il a effectuée. Dans cette famille de méthodes, nous disposons aujourd'hui de procédés plus ou moins sophistiqués.

Martilla et James ont initialement proposé une **double évaluation** pour chaque attribut : la première sur la satisfaction, la seconde sur l'importance. Cette double mesure permet ainsi de construire leur célèbre matrice, en plaçant le niveau de performance en abscisse et le score d'importance en ordonnée, ce qui permet d'identifier quatre zones de critères, des attributs prioritaires (« *concentrate here* ») aux zones de sur-qualité (« *possible overkill* ») notamment. Dans le même ordre d'idée, certains auteurs (Ennew, Reed et Binks, 1993) ont développé des méthodes basées sur la paradigme de confirmation/non-confirmation des attentes et ont proposé de construire un **indice « d'attainment »** qui permet de mesurer la correspondance entre le niveau d'attente et la performance perçue. Cette mesure se situe dans le prolongement des travaux portant sur l'analyse du « *gap* », proposée dans le cadre du modèle SERVQUAL (Parasuraman et al., 1988) Cette technique débouche sur une hiérarchie des attributs prioritaires en fonction du score « *d'attainment* » obtenu, décalage calculé entre attente et satisfaction.

L'**analyse conjointe** est aujourd'hui une méthode reconnue pour déterminer les préférences des consommateurs (Danaher, 1997, DeSarbo et al., 1994) car elle permet d'opérer sur des répondants, des mesures multiples concernant des attributs dont le niveau est rendu variable. L'intérêt de l'analyse conjointe réside principalement dans sa capacité à déterminer l'importance relative de chaque attribut de la satisfaction. Elle permet également d'envisager des relations non linéaires (quadratiques souvent) entre les attributs et la satisfaction globale. Des méthodes plus sophistiquées existent comme par exemple, l'**A.H.P.** ou « procédé de hiérarchie analytique », qui permet également de classer les critères par ordre d'importance aux yeux du consommateur, en se basant sur des comparaisons par paires entre les différents attributs. Ces comparaisons et l'évaluation des écarts d'importance entre les critères conduisent à la construction d'une « super-matrice » qui permet de déterminer les attributs prioritaires pour les répondants (Huber, Fischer et Herrmann, 2000).

Les méthodes induites

Comme son nom l'indique, cette famille de méthodes ne repose pas sur une évaluation directe du poids ou de l'importance de chaque attribut. Il s'agit au contraire de déduire ce poids grâce

à la prise en considération d'une mesure globale de la satisfaction et surtout de la **corrélation** de chaque attribut avec cette note globale, afin de pallier le défaut potentiel des méthodes directes qui ont tendance à favoriser les critères rationnels par rapport aux critères plus « affectifs » (Windal et Desmet, 2000).

La confrontation des évaluations globale et partielle est très intéressante à divers titres. Elle permet tout d'abord de constater l'écart éventuel entre la moyenne de la satisfaction globale et la moyenne des satisfactions partielles. Il n'est guère surprenant de constater que celles-ci sont très rarement équivalentes. On constate presque systématiquement que la note globale est souvent plus faible que la moyenne des notes partielles et que le répondant a une tendance naturelle à être un peu plus sévère dans son évaluation générale. La différence s'explique aisément par les poids relatifs de chaque critère sur l'impression générale, qui ne sont pas identiques : certains items sont plus importants et pèsent donc plus fortement sur l'évaluation globale. En revanche, le fait que cette différence soit généralement au profit de la moyenne des évaluations partielles ne s'explique pas facilement, si ce n'est par le fait que ce sont les points faibles qui semblent avoir un impact relatif plus fort sur la note finale, ce qui tendrait à prouver qu'il convient de bien différencier attributs de satisfaction et attributs d'insatisfaction. Dans cette famille de méthodes également, nous relevons plusieurs options différentes, plus ou moins sophistiquées, qui varient principalement en fonction de la nature de la corrélation calculée.

La méthode la plus simple consiste à étudier le **coefficient de corrélation linéaire simple** entre la satisfaction globale d'une part et la satisfaction partielle sur chaque attribut d'autre part. Bien que relativement réductrice, cette méthode présente un double avantage. Elle évite les problèmes de multi-colinéarité et elle est peu sensible aux données manquantes.

Les auteurs privilégient toutefois la prise en compte simultanée de tous les attributs présumés de la satisfaction, au sein d'une **analyse de régression multiple**. Pour des raisons liées à la multi-colinéarité évoquée plus haut (Danaher et Haddrell, 1996), les travaux récents auraient tendance à préconiser plutôt le recours à une régression multiple sur les moindres carrés partiels, plus connue sous l'appellation de **régression P.L.S.** (Windal et Desmet 2000).

Dans les avancées récentes des méthodes induites, on notera la **différentiation entre attributs de satisfaction et attributs d'insatisfaction**. Il s'agit de concrétiser l'idée majeure selon laquelle la relation entre la performance des attributs et la satisfaction globale est asymétrique (Kano, 1984, Roszkowski, 2003, Matzler et al., 2004). Le poids d'un attribut doit être considéré en deux temps (ou selon deux dimensions) : sa capacité à générer de la

satisfaction en cas de performance d'une part et sa propension à créer de l'insatisfaction en cas de dysfonctionnement d'autre part. Plusieurs méthodes permettent de prendre en compte cette double dimension. L'une d'entre elles consiste à effectuer une régression linéaire multiple, avec une satisfaction globale dichotomique (Matzler et al., 2004). Pour le même contexte, d'autres auteurs préfèrent le recours à la **régression logistique** (Greenland, Coshall et Combe, 2004).

Enfin, pour répondre à ce même objectif, une autre méthode est disponible avec laquelle, à l'issue d'une **analyse factorielle des correspondances** (Llosa, 1997, Lichtle et al., 2002), l'importance d'un attribut sera alors mesurée sur deux dimensions : sa capacité à satisfaire d'un côté et sa capacité à mécontenter de l'autre. Cette méthode permet de classer les différents attributs en 4 classes, en fonction de leurs positions respectives sur l'axe de la satisfaction d'une part et sur l'axe d'insatisfaction d'autre part.

L'effet de halo dans les méthodes induites

Avantages et inconvénients des deux familles de méthodes

Certaines méthodes directes (comme l'analyse conjointe ou l'A.H.P.) sont très précises car elle présentent l'énorme avantage de proposer une comparaison des attributs par paires, ce qui permet *in fine* de générer une hiérarchie vérifiable et quasiment incontestable. Toutefois, on leur reproche une tendance à **privilégier les attributs "rationnels"** au détriment des dimensions plus subjectives ou affectives (Windal et Desmet 2000).

Beaucoup plus simples à mettre en œuvre, les méthodes induites sont de plus en plus répandues et notamment auprès des praticiens. Ces derniers ont tendance à écarter les méthodes directes du fait de l'**extrême lourdeur** de certains protocoles qui -s'ils ne se déroulent pas dans un environnement adapté- ne sont guère opérationnels. Ils demandent par ailleurs au répondant un effort de rationalisation qui semble parfois très artificiel : un consommateur est-il réellement capable par exemple, de hiérarchiser plus d'une dizaine de critères en les comparant un à un ?

Plus souples et plus opérationnelles, les méthodes induites sont très fréquemment utilisées dans les enquêtes de satisfaction menées en pratique par de nombreuses organisations. Elles permettent, par un mécanisme plus **"inconscient"**, une meilleure prise en compte des attributs affectifs. Toutefois, elles peuvent présenter un problème potentiel qui est l'**effet de halo** entre

satisfactions partielles et satisfaction globale. Elles nécessitent donc une bonne compréhension des liens qui peuvent exister réellement entre la satisfaction globale et les satisfactions partielles.

L'effet de halo et son application dans les enquêtes de satisfaction

De nombreuses recherches ont testé par le passé, l'influence de la manière de formuler les questions sur la nature des réponses recueillies. Des travaux antérieurs portant sur l'impact des caractéristiques du questionnaire (Grémy, 1987, Ganassali, 2005a), nous pouvons envisager le modèle suivant, pour décrire les liens entre caractéristiques de l'enquête, contexte de l'étude et qualité des réponses.

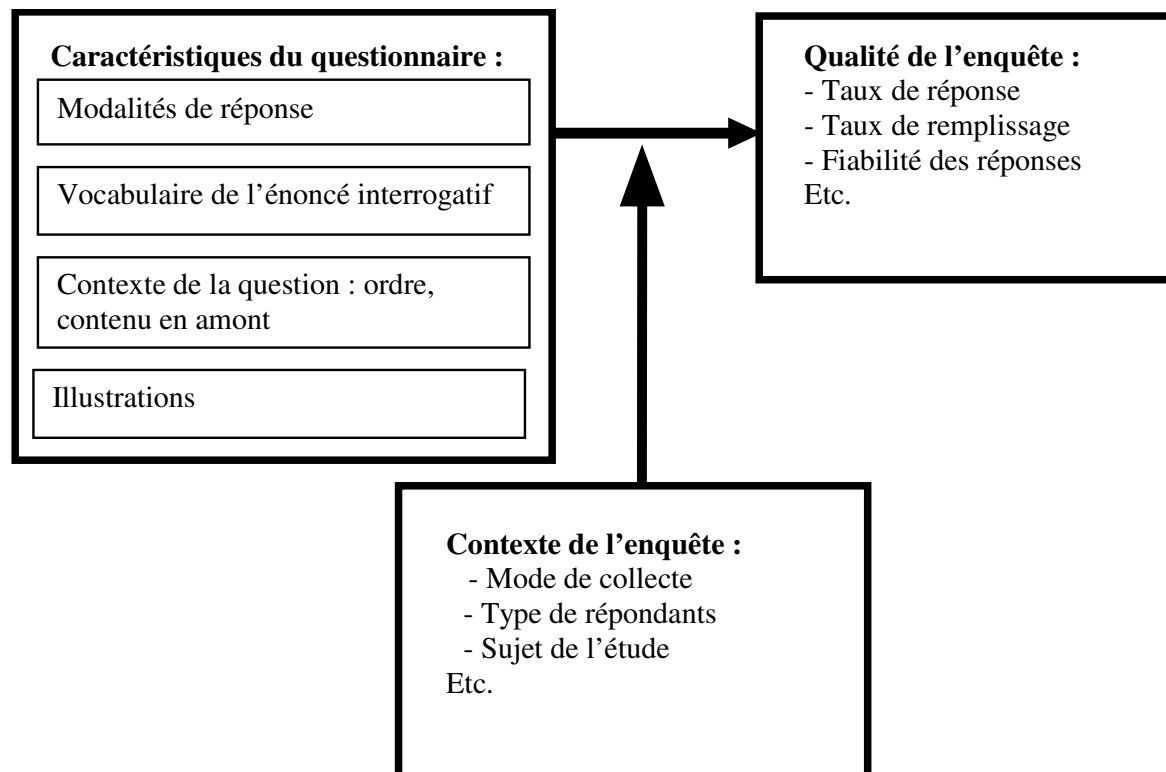


Figure 1 : Modèle général de réflexion

Jusqu'à 2/3 de la variance dans la mesure de la qualité peuvent être expliqués par les caractéristiques de la conception de l'enquête (Andrews, 1984). Parmi les critères les plus influents, on relève le nombre d'échelons utilisés, la longueur de la batterie d'items évalués, la longueur des questions et la position des questions dans le formulaire. Grémy (1987) nous

propose une très instructive revue de la littérature francophone sur le sujet, qu'il met en perspective avec les résultats anglo-saxons dans sa conclusion. Il classe les différents types d'effets abordés dans la littérature en trois catégories :

- les différences dans les modalités de réponse offertes,
- les différentes formes de synonymie ou de paraphrase dans l'énoncé interrogatif,
- le contexte de la question incluant notamment l'ordre des questions et le contenu discursif en amont de la question,

Notre recherche se situe clairement dans la troisième catégorie et se concentre sur une de ses problématiques. Elle s'applique spécifiquement aux enquêtes de satisfaction et à la question-clé de **la position de l'évaluation globale par rapport à celle des évaluations spécifiques des attributs**.

Un des facteurs très importants dans la conception de l'enquête est **l'ordre des questions**. Andrews (1984) constate que la qualité est plus basse pour les questions placées au début et à la fin du questionnaire. Drew (1987) a mis en évidence que le fait de proposer une évaluation de la satisfaction globale en début de questionnaire produisait un effet statistique sur l'évaluation des attributs. Dans une autre étude, menée par téléphone, Drew et Bolton (1991) ont montré que selon la position de l'évaluation globale dans le questionnaire, les résultats pouvaient être assez différents. On constate par exemple, une satisfaction significativement plus forte quand celle-ci est placée en fin d'enquête. Ils concluent que la réponse d'un consommateur sur un item donné peut dépendre de sa position et des items voisins dans le questionnaire.

Depuis Thorndike (1920), l'effet de « **halo** » a été fréquemment étudié dans le passé dans les recherches en psychologie sociale portant spécifiquement sur l'évaluation. Selon la littérature, il peut se définir comme « l'influence d'une évaluation globale sur l'évaluation des attributs spécifiques » (Nisbett et Wilson, 1977). La cause probable du halo est « l'incapacité ou le refus d'un évaluateur à discriminer parmi plusieurs attributs spécifiques » (Cooper, 1981). L'évaluateur a tendance à considérer « l'individu comme un tout » (Thorndike, 1920). Sur le plan théorique, il semble que l'effet de halo puisse s'expliquer par plusieurs mécanismes décrits précédemment dans la littérature (voir Grémy, 1987 pour une synthèse). Nous évoquerons notamment l'effet de « cohérence » qui traduit « la propension à éviter d'éventuelles contradictions entre la réponse donnée et la réponse antérieure ». Par ailleurs,

l'effet de « fatigue » crée une baisse de l'attention et de la réflexion accordée par le répondant aux questions posées en fin d'entretien et peut entraîner une diminution de la discrimination dans les réponses. En conséquence, l'erreur de halo conduit à une inflation artificielle des corrélations entre les attributs évalués. Des travaux récents (Grenwald et Gillmore 1997, Feeley, 2002, Revat, 2005) ont démontré par exemple un effet de halo qui se produit en fonction de la tonalité (positive ou négative) d'une question précédente.

L'effet de halo (Murphy, Jako et Anhalt, 1993, Solomonson et Lance, 1997) peut se diviser en deux composantes : le « vrai halo » et le « halo illusoire ». Le vrai halo reflète la réelle corrélation entre les différents items évalués tandis que le halo illusoire correspond à l'ensemble des erreurs de mesure ou d'observation. C'est donc sur cette deuxième composante que se porte en priorité l'intérêt des chercheurs.

Dans les enquêtes de satisfaction, l'effet de halo peut se définir comme l'influence d'une réponse sur une autre réponse. Suite à une évaluation qui lui est proposée, le répondant va conserver une impression (positive ou négative) et celle-ci va subsister au moment où il devra établir une nouvelle évaluation. Ainsi, l'évaluation d'un attribut donné pourrait être influencé par les évaluations effectuées sur les attributs précédents.

On peut pressentir que l'effet de halo soit actif pour les méthodes d'administration assistées où un enquêteur dévoile progressivement les questions au répondant. Celui-ci devrait être plus diffus (voire inexistant) dans les méthodes auto-administrées ou généralement, le répondant a la possibilité de consulter les questions dans l'ordre qu'il souhaite.

L'impact de la disposition de l'évaluation globale

Dans notre recherche, nous nous intéressons en particulier, à **la position de l'évaluation globale, située soit avant, soit après les évaluations partielles**. Cette différence dans la position de l'évaluation générale pourrait produire des effets réels sur les corrélations entre la note globale et les notes des attributs individuels (hypothèse 1). Nous faisons l'hypothèse que l'évaluation a priori serait plus spontanée : elle résulterait assez naturellement d'une appréciation holistique globale. L'évaluation en fin de questionnaire entraînerait une post-rationalisation de la part du répondant qui serait guidé par une certaine obligation de « cohérence » entre les évaluations précédemment produites sur les attributs spécifiques (Nathan et Tippins, 1990). On peut même envisager un « effet de halo » particulier des

derniers attributs placés juste avant la satisfaction globale (hypothèse 2). Dans le cas d'une évaluation globale postérieure à l'évaluation des attributs, l'effet de halo pourrait produire alors une corrélation artificielle entre la satisfaction globale et certains attributs proches placés avant. En conséquence, dans le cas où celle-ci s'établit avec une méthode induite, la catégorisation des attributs de la satisfaction pourrait en être partiellement faussée.

Cet effet pourrait être spécifique aux modes de collecte assistés (hypothèse 3). Il est possible d'envisager que l'effet de halo puisse se produire beaucoup moins fortement dans les méthodes auto-administrées. En effet, dans les situations d'auto-administration, le répondant peut consulter la batterie d'items dans un ordre différent de celui qui lui est imposé par un éventuel enquêteur, au téléphone ou en face à face. Dans ces conditions, le répondant peut parcourir le questionnaire comme bon lui semble et il peut ainsi envisager sa satisfaction globale avant ses satisfactions partielles, même si elle est située après, sur une autre page ou sur un autre écran. L'effet de cohérence entre les derniers attributs évalués et l'appréciation globale serait donc moins actif, voire quasiment inexistant.

Méthodologie de notre recherche

Afin de vérifier nos hypothèses, nous avons pu avoir accès aux données d'une enquête réalisée chaque année par un grand institut de sondages français pour le compte d'une société dans le secteur de la gestion immobilière. Le questionnaire est administré par téléphone auprès des locataires. Nous avons comparé les résultats de plusieurs vagues annuelles dont les questionnaires étaient totalement identiques à une seule exception près : dans un cas, l'évaluation globale était située avant les évaluations partielles ; dans l'autre la question posée pour mesurer la satisfaction globale était placée après l'évaluation des attributs.

La méthode de l'**échantillon partagé** est préconisée pour ce type de recherche (Grémy, 1987). Elle consiste à « administrer à des sous-échantillons identiques les différentes formes de questions à comparer ».

Pour valider la présence d'un effet de halo, nous avons comparé les résultats obtenus en utilisant 4 différentes méthodes induites de catégorisation des attributs de la satisfaction : la régression linéaire multiple, la régression PLS, la régression logistique sur satisfaction globale dichotomique et le modèle tétra-classes par analyse des correspondances.

En plus d'une évaluation globale, les critères évalués portaient sur la satisfaction des locataires vis-à-vis d'un ensemble de 18 services rendus par le bailleur ou d'éléments tangibles du logement et de l'immeuble. Ils sont listés ici dans l'ordre de leur citation dans le questionnaire ou plus précisément dans l'ordre du script d'entretien téléphonique.

L'accueil des collaborateurs de l'agence
La qualité du contact avec le gardien
La rapidité de réponse du gardien
Le fonctionnement des équipements techniques de l'immeuble
La rapidité de réparation du contrôle d'accès
La rapidité de réparation de l'ascenseur
L'état général des équipements
L'état des fenêtres
L'état des sanitaires et des robinetteries
La fonctionnalité et les aménagements des espaces rangement
L'isolation par rapport aux bruits de l'extérieur
L'isolation par rapport aux bruits intérieurs de l'immeuble
L'aspect général de l'immeuble
L'état des parties communes
L'état de l'ascenseur
L'état des espaces verts
L'entretien ménager effectué dans les parties communes
La qualité de vie dans le quartier

Tableau 1 : Les 18 critères évalués lors de l'enquête de satisfaction

Pour cette enquête qui est un baromètre de satisfaction annuel, nous disposons des données sur 4 années successives, de 2000 à 2003. Les trois premières années, l'item portant sur l'évaluation de la satisfaction globale est placé **avant** les 18 attributs spécifiques, tandis qu'en 2003, pour des raisons liées à des choix méthodologiques de la part de l'institut de sondages, l'évaluation globale est située **après**. Nous avons donc comparé les indicateurs statistiques dans les deux situations, afin de vérifier la production éventuelle d'un effet de halo et son intensité en fonction de la position de l'attribut, et plus précisément de sa proximité de la satisfaction globale. D'une vague à l'autre, la taille et les caractéristiques de l'échantillon sont stables : celui-ci compte environ 2 000 personnes et il est toujours constitué par l'institut de sondages pour être représentatif de la population des locataires de la société. Sur les trois premières années de l'étude, les corrélations entre attributs et satisfaction globale sont stables,

ce qui confirme l'assertion selon laquelle la satisfaction peut varier sensiblement d'une vague à l'autre tandis que l'importance de chaque critère est par nature plus stable dans le temps.

Résultats

Existence d'un effet de halo.

En comparant les différents coefficients obtenus par des analyses réalisées respectivement sur la satisfaction globale placée avant ou placée après, nous constatons des différences significatives. Sur les analyses de régression multiple et PLS (figures 2 et 3), la variation moyenne en valeur absolue des coefficients de régression s'élève à 0,03 pour des coefficients s'échelonnant entre 0 et 0,2, ce qui pourrait traduire un effet de halo important. Pour la régression logistique, la variation moyenne en valeur absolue des coefficients de régression est de 0,15 pour des coefficients compris entre $-0,5$ et $0,5$. Pour l'analyse factorielle des correspondances, les coordonnées des modalités varient en valeur absolue et en moyenne de 0,2 pour des valeurs comprises entre $-0,8$ et $0,4$. Si l'on fait la tentative de le quantifier globalement, en regardant la part de cette variation sur la valeur des coefficients, on pourrait dire que –toutes méthodes confondues- **l'effet de halo peut entraîner une variation moyenne des différents coefficients de l'ordre de 15%**. Le poids relatif de cet effet sur les différents coefficients est pratiquement identique pour les quatre méthodes.

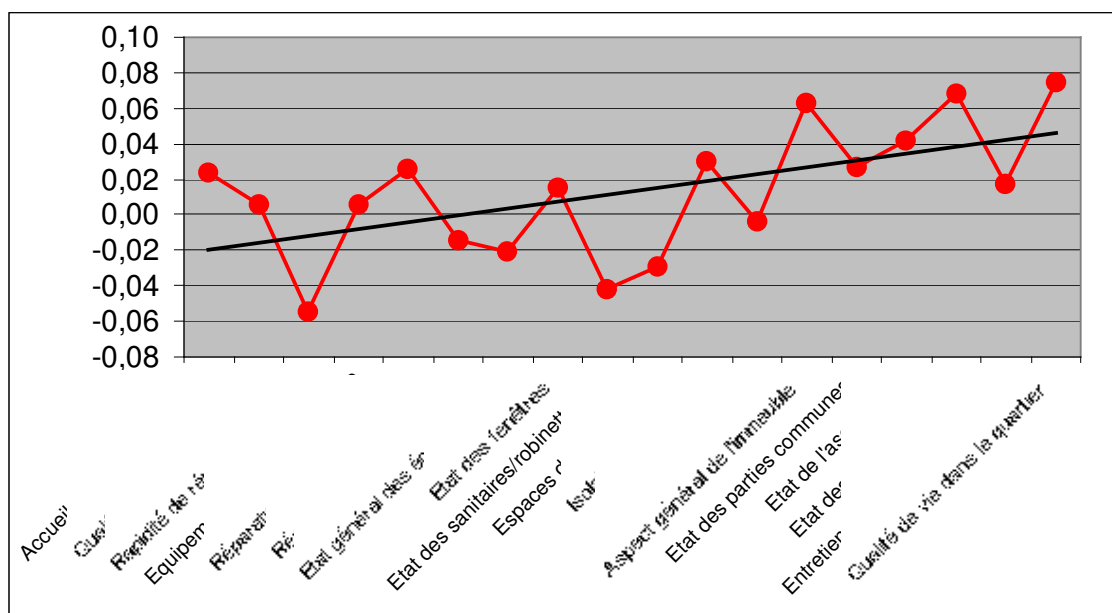


Figure 2 – Halos issus de la régression multiple

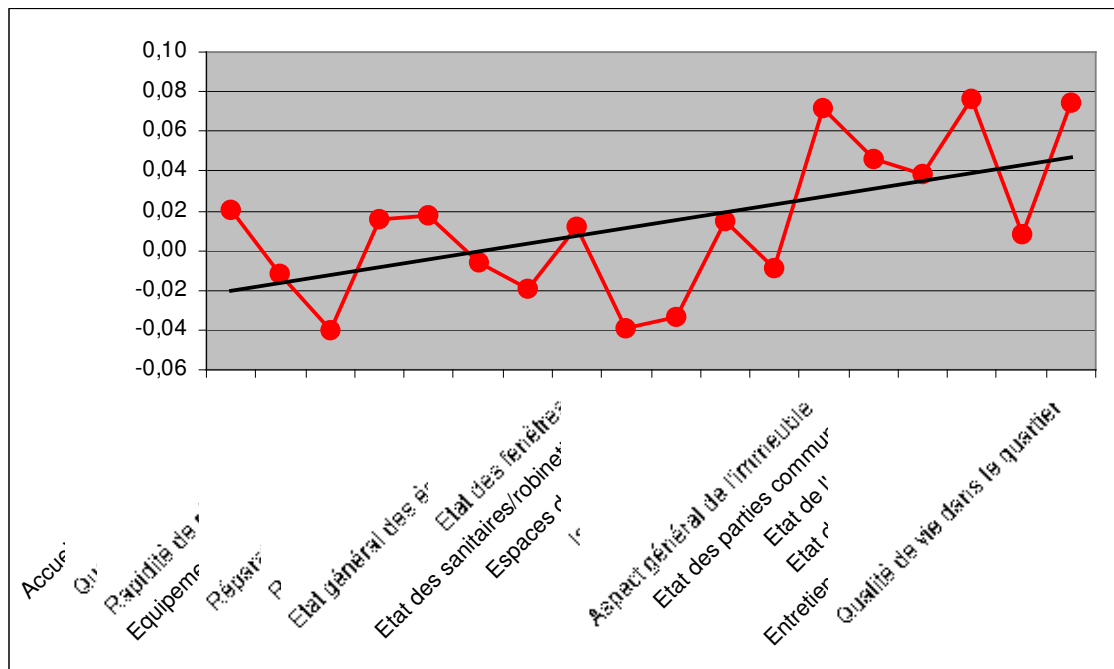


Figure 3 – Halos issus de la régression PLS

Halo et proximité de la satisfaction globale

Sur les graphiques issus des régressions multiples et PLS (voir figures 2 et 3 à nouveau), on remarque que **l'intensité du halo varie en fonction de la position de l'attribut évalué**. Nettement, l'intensité croît à mesure que la proximité de la satisfaction globale augmente. L'effet de halo semble presque nul pour les critères très éloignés tandis que celui-ci atteint des valeurs proches de 0,5 point pour les attributs situés juste en amont de l'évaluation globale. Ces résultats confirmeraient notre deuxième hypothèse selon laquelle l'évaluation globale située à la fin d'une batterie d'items entraînerait une post-rationalisation de la part du répondant, due à un effet de « **cohérence** ». Le répondant apprécie globalement la marque ou l'entreprise, en partie, en fonction des derniers critères qu'ils viennent d'évaluer juste précédemment, créant ainsi une corrélation artificielle entre ces derniers critères et la note globale.

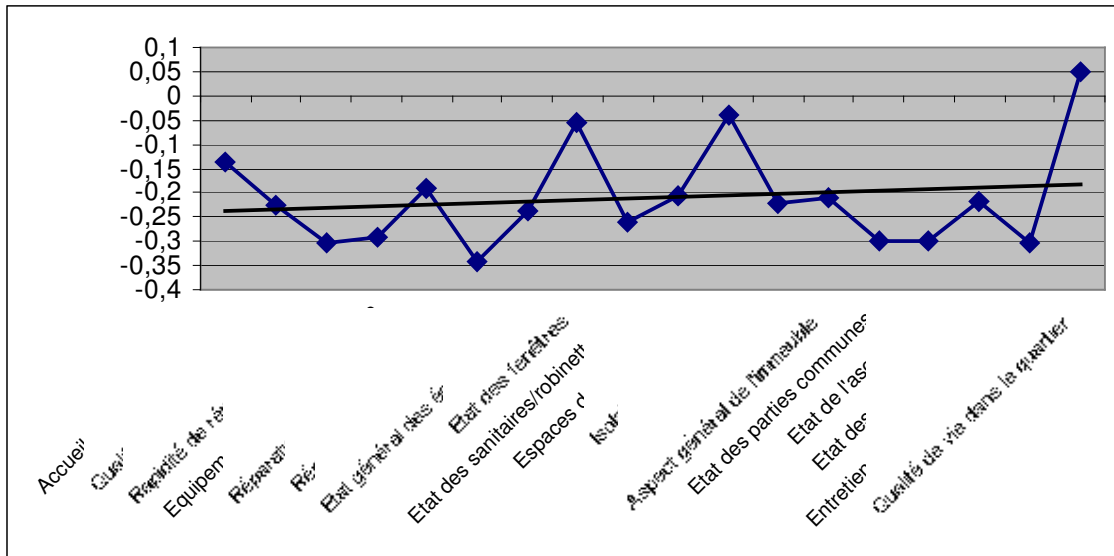


Figure 4 – Halos issus de l'analyse factorielle des correspondances

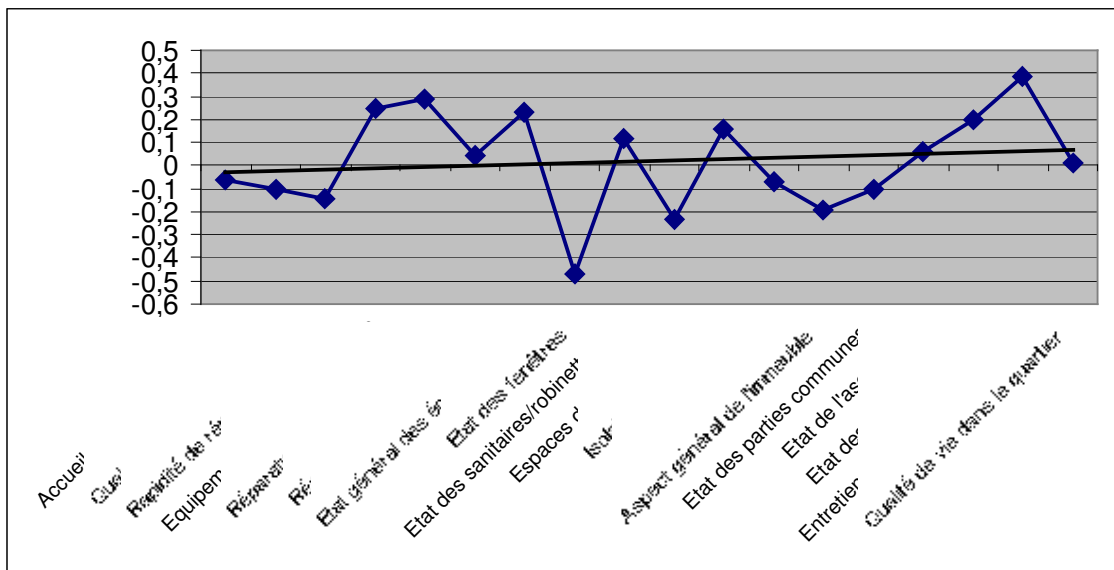


Figure 5 – Halos issus de la régression logistique

Ces effets sont en revanche beaucoup moins marqués dans les résultats que nous observons à l'issue de la régression logistique et de l'analyse factorielle des correspondances. Certes, on constate une tendance légèrement ascendante du halo à mesure que l'on s'approche de la fin de la batterie d'items (voir figures 4 et 5). Toutefois, la pente de cette tendance est très faible et les critères proches de la satisfaction globale sont affectés d'un halo à peine plus important que les plus attributs les plus éloignés.

Si l'on regarde plus dans le détail, on s'aperçoit en fait que cet effet ascendant est plus fort pour l'analyse portant sur la satisfaction (figure 6), tandis qu'il est nul pour l'analyse de l'insatisfaction. On constate cette même divergence pour les 2 méthodes concernées (AFC et régression logistique). Les raisons qui peuvent expliquer cette différence ne sont pas évidentes, elles méritent d'être approfondies, mais il est possible que l'effet de halo - tout comme l'effet de désirabilité sociale (Grémy, 1987) - ne se produise pas de la même manière et selon la même intensité, en fonction de la nature du thème abordé dans l'enquête. Dans notre modèle général (voir figure 1), le thème de l'étude est une variable modératrice des effets de la construction du formulaire sur la qualité des réponses. Ces résultats ne sont pas incohérents mais ils devront être étudiés plus en profondeur dans le cadre d'autres travaux.

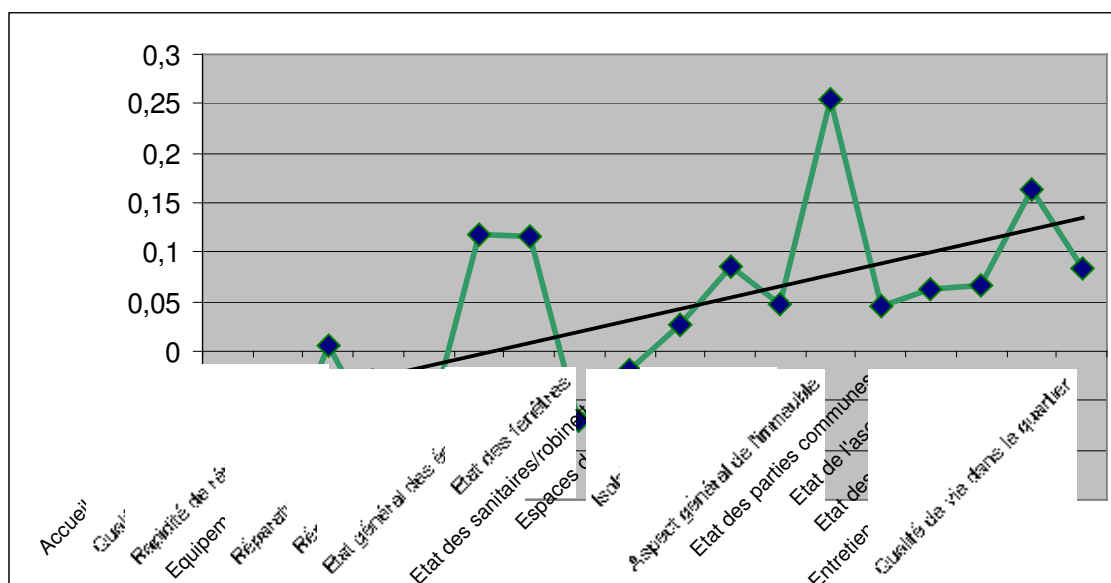


Figure 6 – Halos de la régression logistique pour la satisfaction positive

Halo et mode de collecte

Une première recherche conduite sur le sujet (Ganassali, 2005b) confirmerait cette troisième hypothèse. Une enquête menée sur Internet portant sur l'appréciation des chaînes de télévision hertziennes a été mise en oeuvre selon le même protocole que celui que nous avons utilisé dans l'étude téléphonique précédemment évoquée. Simultanément, un formulaire a été envoyé avec la satisfaction globale située en amont de 10 attributs spécifiques à la moitié des contacts, un autre formulaire avec l'évaluation globale située après, a été diffusé à l'autre moitié de la population-cible. 500 réponses en tout ont été recueillies. Issue d'une régression

multiple, les résultats montrent clairement que le halo n'est pas ascendant à mesure que l'on s'approche de la satisfaction globale et que la courbe de tendance est plate.

Ces premières conclusions sont à prendre cependant avec beaucoup de prudence car la nature de cette deuxième étude diffère assez nettement de la première tant sur le sujet (télévision contre logement), sur le nombre d'attributs évalués (10 contre 18) que sur la taille globale de l'échantillon (500 contre 7 500). Ils constituent néanmoins une première piste de réflexion qu'il nous semble intéressant de mentionner et de poursuivre.

Plusieurs concepts de satisfaction globale

Selon de nombreux auteurs, la "**performance globale**" peut être considérée selon plusieurs approches différentes (voir Lance et Woehr, 1986). Dans une première conception, les évaluations de la performance seraient le résultat indirect des vrais « scores » de l'entité évaluée (qui se manifestent par des faits et des comportements) associés aux biais dus à l'évaluateur. Dans cette conception, la performance globale réelle existe seulement dans l'esprit de l'évaluateur. Selon une autre définition, l'évaluation globale réelle serait un construit exogène de plus haut niveau qui causerait de la covariance entre les évaluations réelles des dimensions spécifiques. Enfin, la performance globale réelle serait une seule et unique composante latente, déterminée par une combinaison linéaire des vraies évaluations partielles.

On voit ici que dans la deuxième définition, la performance globale est considérée comme agissant *a priori*, en amont des évaluations partielles *tandis que dans la troisième, la performance globale est une résultante a posteriori* des évaluations produites sur un ensemble d'attributs spécifiques. Drew et Bolton (1991) suggèrent d'ailleurs qu'il existe deux formes de satisfaction globale : la première qui reflète une évaluation *a priori* de type « *top-of-mind* » et la seconde qui incorpore les attributs appréciés préalablement dans le questionnaire. On pourrait donc envisager qu'en plaçant la satisfaction globale en amont ou en aval des attributs, on ne mesure pas exactement le même concept. Le problème se situe néanmoins dans l'erreur possible que l'on commet, si l'on considère une satisfaction globale évaluée ultérieurement aux attributs comme une performance globale *a priori*. C'est pourtant ce que l'on fait implicitement, quand on utilise comme indicateur d'importance la corrélation des attributs spécifiques avec la note globale.

Conclusion et voies de recherche futures

Si l'on utilise une méthode induite de catégorisation des attributs de la satisfaction, il est plus juste (tant sur le plan théorique que statistique) d'opter pour une évaluation globale mesurée spontanément et antérieurement aux évaluations des attributs spécifiques. Celle-ci correspond bien à la performance globale « *a priori* » que l'on souhaite appréhender dans ce type de démarches.

Dans le cas contraire d'une évaluation globale postérieure, on ne mesure plus une performance de même nature. Cette dernière sera plutôt une combinaison des appréciations partielles, faussée de plus par un effet de halo qui se manifeste entre les derniers critères évalués et la note globale. Il est donc très hasardeux de se fier aux scores d'importance induits, issus de corrélations enregistrées entre les attributs et une satisfaction globale postérieure.

Ces effets semblent bien confirmés avec des méthodes de collecte assistées (comme l'enquête par téléphone) mais ils restent à démontrer pour les méthodes auto-administrées. Il serait donc intéressant de pouvoir mener une enquête de satisfaction qui serait déployée simultanément au téléphone et sur Internet, en faisant varier la position de la satisfaction globale, afin de confronter les halos enregistrés dans ces deux contextes différents de recueil de données. Par ailleurs, il sera utile de tester d'éventuelles différences en fonction de la nature du thème de l'enquête : sujet plus ou moins impliquant ou plus ou moins sensible.

Références bibliographiques

Andrews F. M. (1984), Construct Validity and Error Components of Survey Measures: A Structural Modeling Approach, *Public Opinion Quarterly*, volume 48, page 409

Bartikowski B. & Llosa S. (2004), Customer Satisfaction Measurement: Comparing Four Methods of Attribute Categorizations, *Service Industries Journal*, 25, 3.

Chu, R. (2002), Stated-importance versus Derived-importance Customer Satisfaction Measurement, *Journal of Services Marketing*, 16(4), 285-301.

Cooper, W. H. (1981), Conceptual Similarity as a Source of Illusory Halo in Job Performance Ratings, *Journal of Applied Psychology* 66; 302-307.

Danaher P. J. (1997), Using Conjoint Analysis to Determine the Relative Importance of Service Attributes Measured in Customer Satisfaction Surveys, *Journal of Retailing*, 73, 2, 235-260.

Danaher, P. J., & Haddrell, V. (1996), A comparison of question scales used for measuring customer satisfaction. *International Journal of Service Industry Management*, 17(4), 4-26.

DeSarbo, W.S., Huff, L., Rolandelli, M.M. & Choi, J. (1994), Chapter 9: On the measurement of perceived service quality: A conjoint analysis approach. In R.T. Rust & R.L. Oliver (Eds.), *Service quality: new directions in theory and practice* (pp. 201-222). Thousand Oaks, CA: SAGE Publications.

Drew, J. H. (1987), "Modeling Halo Effects in Surveys for Low-Salience Services," *Proceedings of the Social Statistics Section*, American Statistical Association. Washington, Amer. Stat. Assoc.

Drew, J.H. & Bolton, R.N. (1991), The structure of customer satisfaction: effects of survey measurement, *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behaviour*, Vol. 4 pp.21-31.

Ennew, C. T., Reed G. V. & Binks M. R. (1993), Importance-Performance Analysis and the Measurement of Service Quality, *European Journal of Marketing*, 27 (2), 59-70.

Feeley T. (2002), Evidence of Halo Effects in Student Evaluations of Communication Instruction, *Communication Education*, Vol 5, No. 3 July 2002, pp 225-236.

Ganassali S. (2005a), Les caractéristiques rédactionnelles du questionnaire : fondements et pratiques, *Colloque francophone sur les Sondages*, Université Laval, Québec.

Ganassali S. (2005b), L'effet de halo dans les enquêtes par Internet, *Note de Recherche I.R.E.G.E.*, Université de Savoie.

Greenland S., Coshall J. & Combe I. (2004), Evaluating service quality & customer satisfaction in emerging markets, *Working paper*, London Metropolitan University.

Greenwald A. & Gillmore G. (1997), Grading Leniency Is a Removable Contaminant of Student Ratings, *American Psychologist*, 52(11), pp. 1209-1217.

Grémy J.-P. (1987), Les expériences françaises sur la formulation des questions d'enquête. Résultats d'un premier inventaire, *Revue Française de Sociologie*, XXVIII-4 (octobre-décembre), 567-599.

Huber F., Fischer M. & Herrmann A. (2000), Supermatrix-Analysis as a Method of Measuring Interdependent Relative Importance Weights in Customer Satisfaction Research, in Hoch, S. und Meyer, R. (Hrsg.), *Advances in Consumer Research*, Vol. 27, 92-99.

Kano, N., (1984), Attractive quality and must be quality, Hinshitsu , *The Journal of the Japanese Society for Quality Control*. 14 (2), pp 147-156.

Lance, C. E. & Woehr, D. J. (1986), Statistical control of halo: clarification from two cognitive models of the performance appraisal process, *Journal of Applied Psychology*, 71, 4, 679 - 85.

Lichtle M.-C., Llosa S. & Plichon V., (2002), La contribution des différents éléments d'une grande surface alimentaire à la satisfaction du client, *Recherche et Applications en Marketing*, vol.17, n°4, 23-34.

Llosa S. (1997), L'analyse de la contribution des éléments du service à la satisfaction : Un modèle « tétracasse », *Décisions Marketing*, 10, 81-88.

Martilla, J. A., & James, J. C. (1977). Importance– performance analyses. *Journal of Marketing*, 41(1), 77– 79.

Matzler K., Bailom F., Hinterhuber H., Renzl B. & Pichler J. (2004), The asymmetric relationship between attribute-level performance and overall customer satisfaction: A reconsideration of the Importance-Performance Analysis, *Industrial Marketing Management*, Vol. 33, May, pp. 271-277

Murphy, K.R., Jako, R.A., & Anhalt, R.L. (1993). Nature and consequences of halo error: A critical analysis, *Journal of Applied Psychology* 78(2), 218-225.

Nathan, B. R., & Tippins, N. (1990). The consequences of halo error in performance ratings: A field study of the moderating effect of halo test validation results, *Journal of Applied Psychology*, 75 (3), 290-296.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, 35, 250-256.

Parasuraman A., Zeithaml V. and Berry L. (1988), "SERVQUAL: A Multiple-Item Scale for Measuring Customer Perceptions of Service Quality," *Journal of Retailing*, Spring, pp. 12-40.

Ray D. (2004), Mesure de la satisfaction clients : pratiques et besoins actuels des entreprises, positionnement des offres et principaux enjeux méthodologiques, *Revue Française de Marketing*, 197, 7-21.

Revat R. (2005), La rédaction du questionnaire : quelques expériences, *Colloque francophone sur les Sondages*, Université Laval, Québec.

Roszkowski M. J. (2003), The Nature of the Importance-Satisfaction Relationship in Ratings: Evidence from the Normative Data of the Noel-Levitz Student Satisfaction Inventory, *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, Volume 16, 211-221.

Solomonson, A. L. & Lance, C. E. (1997), Examination of the relationship between true halo and halo error in performance ratings, *Journal of Applied Psychology*, 82, 5, 665 - 74.

Thorndike, E. L. (1920), "Constant Error in Psychological Ratings." *Journal of Applied Psychology* 4: 25-29.

Winda P. & Desmet P. (2000), Les méthodes de mesure de l'importance des critères de satisfaction, *Revue Française de Marketing*, 179-180, 205-220.