

Propriétés de mesure des enquêtes en ligne

Roger Tourangeau¹

Résumé

Les enquêtes en ligne comportent des lacunes graves du point de vue de la représentativité, mais elles semblent comporter certaines bonnes propriétés de mesure. Cet exposé est axé sur les caractéristiques générales des enquêtes en ligne qui influent sur la qualité des données, et plus particulièrement le fait qu'elles sont principalement visuelles. En outre, il examine l'efficacité des enquêtes en ligne comme forme d'autoadministration de questionnaires. Un certain nombre d'expériences ont permis de comparer les enquêtes en ligne avec d'autres modes de collecte de données. Une méta-analyse de ces études montre que les enquêtes en ligne conservent les avantages des formes traditionnelles d'enquêtes autoadministrées; de façon plus particulière, elles réduisent le biais de désirabilité sociale par rapport aux questions administrées par un intervieweur. En conclusion, j'examine certains progrès futurs probables des enquêtes en ligne, à savoir l'intégration d'avatars comme « intervieweurs virtuels » et l'utilisation croissante de dispositifs mobiles (comme les tablettes électroniques et les téléphones intelligents) pour accéder aux enquêtes en ligne et y répondre.

Mots clés : Erreur de mesure; effets de mode; caractéristiques visuelles des enquêtes.

1. Introduction

Les enquêtes en ligne comportent plusieurs caractéristiques importantes qui peuvent réduire l'erreur de mesure par rapport à d'autres méthodes de collecte des données :

- 1) Le principal mode de présentation des questions est visuel, ce qui permet l'intégration de photographies, de vidéoclips, etc.;
- 2) Les questions sont administrées par l'ordinateur plutôt que par un intervieweur, une caractéristique qui peut réduire le biais de désirabilité sociale dans les réponses et qui comporte d'autres avantages aussi;
- 3) L'enquête peut être interactive, en fournissant divers types d'aide aux répondants et en dirigeant les répondants vers les questions appropriées;
- 4) Le répondant contrôle le rythme des questions de l'enquête et peut facilement relire les questions, des caractéristiques qui peuvent réduire le fardeau cognitif de la réponse.

Dans cet exposé, je mets l'accent sur la première et la troisième de ces caractéristiques.

2. Heuristique visuelle pour les enquêtes en ligne

Parmi les caractéristiques potentielles souhaitables des enquêtes en ligne figure leur capacité de présenter de l'information visuelle, y compris des images fixes ou des vidéos. Les images sont susceptibles d'attirer l'attention des répondants et, lorsqu'elles sont présentées en parallèle avec les questions, elles sont susceptibles d'influencer l'interprétation des questions par les répondants. Couper, Tourangeau et Kenyon (2004) ont mené une expérience

¹Roger Tourangeau, Westat, 1600, Research Boulevard, Rockville, Maryland, 20905, États-Unis, RogerTourangeau@Westat.com

dans laquelle ils ont fait varier le contenu des photographies accompagnant chacune des six questions d'enquête. Dans les questions, on demandait aux répondants à quelle fréquence ils avaient fait diverses choses, comme assister à des événements sportifs ou faire un voyage d'une nuitée, au cours de l'année précédente. Les photographies présentées avec les questions décrivaient un cas de la catégorie d'intérêt. Les photos ont été choisies pour représenter des exemples à faible fréquence ou à fréquence élevée de la catégorie cible (par exemple, une partie du baseball majeur ou une partie de « Little League » (partie d'enfants)). Les répondants à qui on a présenté des cas à fréquence élevée de la catégorie cible ont eu tendance à déclarer des fréquences comportementales plus élevées que ceux à qui on a présenté des cas à faible fréquence. Une étude postérieure de Couper et de ses collaborateurs a aussi démontré les effets du contexte visuel, cette fois sur l'évaluation par le répondant de sa santé (Couper, Conrad et Tourangeau, 2007).

L'espacement et d'autres indices visuels peuvent aussi influencer la façon dont les répondants interprètent leurs tâches ou la signification prévue des échelles de réponse. Selon Tourangeau et ses collègues (Tourangeau, Couper et Conrad, 2004, 2007), les répondants suivent quatre heuristiques d'interprétation pour attribuer une signification aux échelles de réponse des enquêtes en ligne. Les quatre heuristiques sont les suivantes :

- milieu signifie typique ou central;
- gauche et sommet signifient premier;
- proche signifie apparenté; et
- semblable (en apparence) signifie proche (en signification).

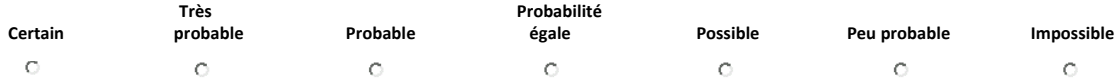
Selon la première heuristique, le point médian visuel d'une échelle joue un rôle important dans l'établissement de la signification des points de l'échelle. Le point médian visuel d'une échelle bipolaire représentera donc, selon l'heuristique, le point médian conceptuel de l'aspect du jugement, c'est-à-dire le point neutre de l'échelle. Lorsque la dimension sous-jacente est unipolaire, les répondants peuvent présumer que le point médian visuel représente la valeur la plus typique (la médiane ou le mode de la population). La deuxième heuristique (« gauche et sommet signifient premier ») correspond à l'attente que les options de réponse suivent un certain ordre logique. Lorsque les options sont affichées horizontalement, les répondants s'attendent à ce que l'option la plus à gauche représente un extrême et l'option la plus à droite l'autre extrême, et à ce que les options qui restent suivent un ordre conceptuel de gauche à droite. Lorsque les options sont affichées verticalement, le répondant s'attend à ce que les options supérieures et inférieures représentent les deux extrêmes, et à ce que les options qui restent suivent un ordre du sommet à la base. Tourangeau et ses collègues (Tourangeau et coll., 2004, Expérience 4) ont trouvé que les répondants prenaient beaucoup plus de temps pour répondre aux questions lorsque l'ordre des options n'était pas conforme à ces attentes que lorsqu'il était conforme.

L'heuristique « proche signifie apparenté » a trait à la tendance pour les répondants d'inférer un lien conceptuel entre deux questions sur la base de leur proximité physique. Dans une étude examinant cette heuristique, Tourangeau, Couper et Conrad (2004, Expérience 6) ont comparé trois méthodes de présentation d'une série de questions d'attitude connexes. Ils ont présenté les questions : a) dans une grille sur un écran; b) dans deux grilles sur deux écrans différents; ou c) comme questions individuelles sur des écrans adjacents. Ils ont prédit (et déterminé) que les corrélations les plus fortes entre les questions se produisaient lorsque toutes étaient présentées dans la même grille, et les plus faibles lorsque chaque question était affichée sur un écran distinct (voir aussi Couper, Traugott et Lamias, 2001). Lorsque les questions étaient placées dans une grille ensemble, les répondants semblaient s'attendre à ce qu'elles aient une signification similaire et ils avaient tendance à leur donner des réponses plus uniforme. La quatrième heuristique (« semblable (en apparence) signifie proche (en signification) ») a trait à la tendance pour les répondants d'inférer une similitude conceptuelle entre deux questions ou deux options de réponse selon la similitude de leur apparence. Par exemple, lorsque les deux extrémités de l'échelle de réponses sont dans la même nuance de couleur, les répondants peuvent inférer que les deux extrémités sont plus près conceptuellement que lorsque les deux extrémités sont dans des nuances de couleurs différentes. Selon cette inférence, les réponses des répondants peuvent changer (Tourangeau et coll., 2007).

La figure 2-1 ci-après fournit une illustration des répercussions de la première heuristique. Dans la version de l'échelle du haut, le point médian conceptuel (« probabilité égale ») coïncide avec le point médian visuel. Dans la version du bas, le point médian conceptuel se situe à droite du point médian visuel. Nous (Tourangeau, Couper et Conrad, 2004) avons mené une expérience pour comparer les deux versions. Dans la version de l'échelle du bas, les répondants étaient significativement plus susceptibles de choisir un des points de l'échelle à faible probabilité (« possible », « peu probable », ou « impossible »).

Figure 2-1
Déplacement du point médian conceptuel d'une échelle

Au cours de la prochaine année, quelle est la probabilité que vous soyez malade et que vous deviez demeurer au lit pendant toute une journée ou plus longtemps?



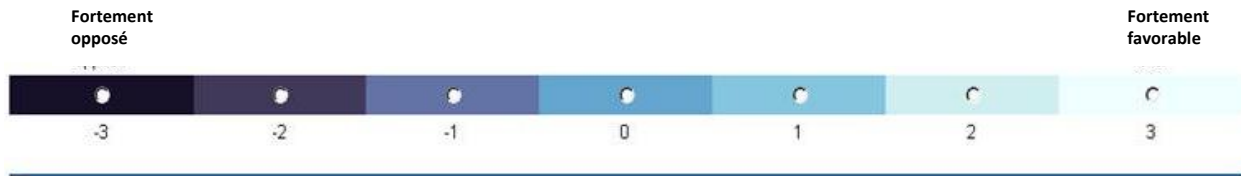
Au cours de la prochaine année, quelle est la probabilité que vous soyez malade et que vous deviez demeurer au lit pendant toute une journée ou plus longtemps?



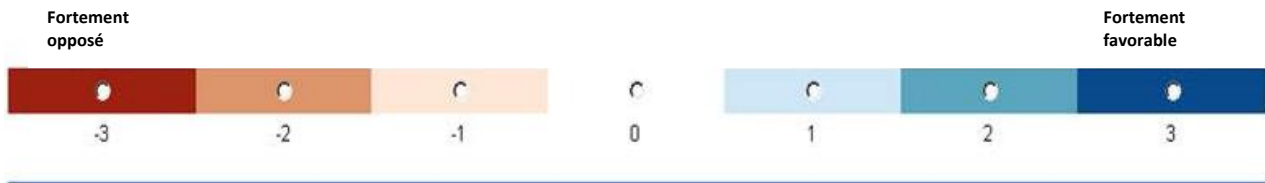
La figure 2-2 ci-après illustre l'heuristique « semblable (en apparence) signifie proche (en signification) ». Dans les deux expériences, les répondants ont reçu les deux échelles, qui étaient ombrées dans la même nuance de couleur (comme dans la partie du haut de la figure) ou dans des nuances différentes aux deux extrémités de l'échelle.

Figure 2-2
Nuance d'une couleur versus nuance de deux couleurs d'une échelle de points

Dans quelle mesure favorisez-vous d'éviter « la malbouffe » ou vous y opposez-vous?



Dans quelle mesure favorisez-vous d'éviter « la malbouffe » ou vous y opposez-vous?

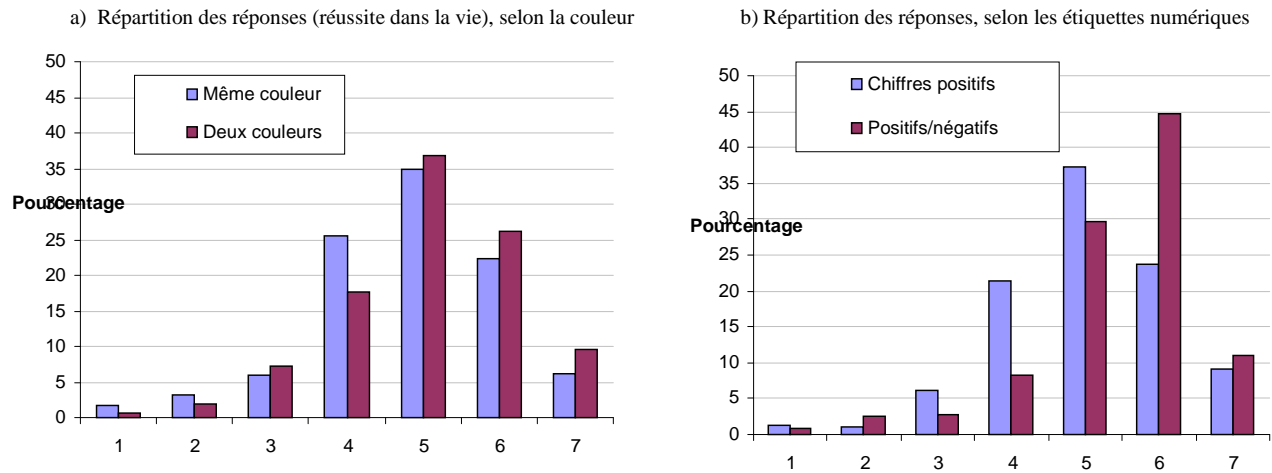


Dans l'expérience, on a aussi fait varier l'étiquetage des points de l'échelle. Certains répondants ont reçu des étiquettes verbales uniquement pour les points des extrémités de l'échelle, comme dans les exemples de la figure. En outre, à l'intérieur de cette condition, nous avons fait varier la façon dont les points de l'échelle étaient numérotés. Pour certains répondants, les points de l'échelle étaient numérotés de 1 à 7; pour d'autres, les chiffres allaient de -3 à +3 (encore une fois comme l'illustre la figure); pour un troisième groupe, les points de l'échelle n'étaient pas numérotés. Un quatrième groupe a reçu des échelles entièrement étiquetées, c'est-à-dire que chaque point de l'échelle comportait une étiquette verbale. Nous avons prédit que les répondants utiliseraient la couleur comme un indice de la signification prévue des points de l'échelle, tout comme ils utilisaient les chiffres (comme le démontrent

Schwarz et coll., 1991). La figure 2-3 montre les résultats de l'expérience. Dans la version en deux couleurs de l'échelle de réponses, les réponses sont poussées vers la droite (panel a), mais dans une moins large mesure que dans l'échelle utilisant à la fois des chiffres positifs et négatifs (panel b).

Figure 2-3

Répartitions des réponses, selon les nuances des points de l'échelle (graphique de gauche) et les étiquettes numériques (graphique de droite).



3. Enquêtes en ligne comme forme d'autoadministration des questionnaires

L'une des différences les plus importantes entre les modes d'enquête est la présence ou non d'un intervieweur pour poser les questions et consigner les réponses. Un ensemble substantiel de preuves montrent que les répondants sont plus prêts à faire des révélations potentiellement embarrassantes dans des enquêtes lorsqu'un intervieweur n'est *pas* présent et qu'ils remplissent plutôt des questionnaires sur papier ou répondent à des questions qui s'affichent à l'ordinateur (Tourangeau et Yan, 2007). L'autoadministration, que ce soit par ordinateur ou au moyen d'un questionnaire sur papier, permet d'obtenir des réponses plus précises aux questions de nature délicate sur des sujets comme la consommation de drogues illicites, la consommation d'alcool et le vote. Évidemment, un aspect clé des enquêtes en ligne est qu'elles ne font intervenir aucun intervieweur.

Une étude de Kreuter, Presser et Tourangeau (2007) démontre cet avantage d'une enquête en ligne par rapport à d'autres méthodes de collecte des données. Kreuter et ses collègues ont comparé les erreurs de déclaration selon trois modes de collecte des données, ITAO, RVI et enquête en ligne. Les investigateurs ont été en mesure de valider les réponses à certaines des questions d'enquête par rapport à des données de dossiers externes. Pour les quatre comportements socialement indésirables qu'ils ont examinés (des choses comme avoir une note insatisfaisante en classe ou faire l'objet d'une probation académique), l'enquête en ligne a produit le taux le plus faible d'erreurs de déclaration, plus faible que pour l'ITAO (qui nécessite l'administration du questionnaire par un intervieweur) et la RVI (qui ne fait pas intervenir d'intervieweur). Chang et Krosnick (2009) ont aussi examiné les répercussions de l'administration d'une enquête en ligne sur le taux de réponses socialement désirables. Ils ont déterminé que les répondants de race blanche étaient plus susceptibles de dire que le gouvernement devrait fournir de l'aide aux Américains de race noire lorsque les questions étaient administrées par des intervieweurs au téléphone que lorsque les mêmes questions étaient administrées dans une enquête en ligne; ils interprètent ces résultats comme signifiant que les répondants sont plus portés à fournir des réponses véridiques (mais socialement indésirables) à une enquête en ligne. Certaines enquêtes en ligne comportent des « avatars » (représentation à l'écran du chercheur ou d'un intervieweur) ou d'autres touches « humanisantes »; jusqu'à maintenant, ces caractéristiques ne semblent pas augmenter les biais de désirabilité sociale (Tourangeau, Couper et Steiger, 2003). Toutefois, ces améliorations au texte devraient probablement être évitées lorsque la désirabilité sociale pourrait être une source de préoccupation. En

outre, les interfaces « adaptées à chaque sexe » (par exemple, un questionnaire en ligne qui affiche une photographie de l'investigatrice féminine) peuvent influencer les réponses aux questions concernant des sujets liés au sexe (par exemple, des questions concernant les femmes sur le marché du travail; Tourangeau et coll., 2003; voir aussi Fuchs, 2009).

Tourangeau, Conrad et Couper (2013) ont procédé à une méta-analyse des résultats d'un certain nombre d'expériences comparant les enquêtes en ligne avec d'autres modes de collecte des données. Au total, les études incluaient 223 comparaisons de mode, dont 160 faisaient intervenir des questionnaires en ligne et sur papier. Le tableau 3-1 montre les principaux résultats de leur méta-analyse; il présente les tailles d'effet moyen (généralement le logarithme des rapports de cotes) pour chacune des études.

Tableau 3-1.
Tailles des effets selon l'étude et la comparaison des modes

Comparaison	Étude	Tailles d'échantillon	Taille de l'effet moyen	Erreur type	
En ligne et poste/papier	Bälter et coll. (2005)	Poste : 188 En ligne : 295	0,054	0,309	
	Bason (2000)	Poste : 204 En ligne : 115	-0,168	0,129	
	Bates et Cox (2008)	Poste : 73 En ligne : 64	-0,014	0,180	
	Eaton et coll. (2010)	Poste : 1 729 En ligne : 3 498	0,070	0,012	
	Denscombe (2006)	Poste : 267 En ligne : 69	-0,256	0,120	
	Knapp et Kirk (2003)	Poste : 174 En ligne : 57	-0,077	0,119	
	Link et Mokdad (2005a, 2005b)	Poste : 836; 804-820 En ligne : 1 143; 948-1 139	0,068	0,039	
	McCabe et coll. (2002, 2006) McCabe (2004)	Poste : 1 412 En ligne : 2 194	0,006	0,019	
En ligne et téléphone	Bason (2000)	Téléphone : 161 En ligne : 115	-0,503	0,132	
	Chang et Krosnick (2007)	Téléphone avant l'élection : 1 456 En ligne avant l'élection : HI : 2 313 KN : 4 914 Téléphone après l'élection : 1 206 En ligne après l'élection : HI : 1 040 KN : 3 408	0,172	0,035	
	Knapp et Kirk (2003)	Téléphone : 121 En ligne : 57	0,193	0,126	
	Kreuter et coll. (2009)	Téléphone : 320 En ligne : 363	0,157	0,060	
	Link et Mokdad (2005a, 2005b)	Téléphone : 2 072; 2 066-2 070 En ligne : 1 143; 948-1 139	0,026	0,031	
	En ligne et RVI	Bason (2000)	RVI : 128 En ligne : 115	0,108	0,143
		Kreuter et coll. (2009)	RVI : 320 En ligne : 363	0,081	0,060

Nota : Adapté de Tourangeau, Conrad et Couper (2013). HI désigne le panel interactif Harris, et KN le panel Knowledge Network; RVI désigne la réponse vocale interactive.

Tourangeau et ses coauteurs ont tiré deux conclusions principales des résultats. Tout d'abord, la collecte de données en ligne semble produire davantage de déclarations de données délicates que les enquêtes téléphoniques administrées par un intervieweur (Chang et Krosnick, 2007; Kreuter, Presser et Tourangeau, 2008; Link et Mokdad,

2005a, 2005b; mais voir Bason, 2000, pour une exception apparente). Parmi les six études comparant les deux modes, la taille de l'effet global était de 0,088, ce qui n'est pas très significatif ($t=1,69$, $df=7$), en raison en partie de l'étude mentionnée par Bason. Si on laisse cette étude de côté, la taille de l'effet moyen pour les comparaisons téléphone-en ligne passe à ,105 (avec une erreur type de ,052). Au moins une étude (Kreuter et coll., 2008) montre que l'augmentation des données déclarées correspond à une augmentation de l'exactitude. Ces résultats correspondent aux analyses antérieures des avantages de l'autoadministration des questionnaires pour recueillir des données potentiellement embarrassantes (Tourangeau et Yan, 2007).

En deuxième lieu, l'administration en ligne semble offrir un petit avantage seulement au chapitre de la déclaration par rapport à d'autres formes d'autoadministration. Par exemple, la taille de l'effet moyen global pour les comparaisons en ligne-papier était de 0,030, avec une erreur type de 0,023.

Mise à part la réduction des biais de désirabilité sociale et des effets de l'intervieweur, les enquêtes en ligne comportent tous les avantages habituels de l'automatisation, y compris la capacité d'adapter les questions sur la base des données chargées au préalable ou des réponses données à des questions précédentes, les sauts de questions et l'acheminement automatisés, les vérifications de contrôle, qui font en sorte que les réponses se situent à l'intérieur d'une fourchette prédéterminée, la randomisation de l'ordre des questions ou des options de réponse, etc. Grâce à ces caractéristiques, il est possible d'administrer des questionnaires très complexes dans les enquêtes en ligne.

4. Conclusions et orientations pour l'avenir

Les enquêtes en ligne continuent d'évoluer, et certaines des conclusions présentées ici devront probablement être modifiées dans quelques années. La couverture de la population en général continuera presque certainement à augmenter, et les chercheurs trouveront sans doute des façons d'améliorer les taux de réponse aux enquêtes en ligne. Toutefois, il semble peu probable que la couverture des enquêtes en ligne s'approche de la couverture des enquêtes téléphoniques ou en personne bientôt. Les problèmes de couverture des enquêtes en ligne sont compliqués par les difficultés de la sélection d'un échantillon représentatif d'utilisateurs du Web, ce qui a forcé de nombreuses enquêtes en ligne à dépendre d'échantillons de volontaires auto-sélectionnés. Compte tenu de la gravité de ces problèmes, il est peu surprenant que les résultats des enquêtes en ligne ne correspondent pas bien aux chiffres de la population en général ou que les rajustements de pondération ne permettent pas de supprimer tous les écarts.

Toutefois, lorsque la représentativité des répondants ne constitue pas une exigence importante, les enquêtes en ligne ont beaucoup à offrir. Elles peuvent être beaucoup moins coûteuses que les autres formes de collecte de données et, comparativement à d'autres types d'échantillons non probabilistes, les panels en ligne peuvent fournir un niveau de diversité parmi les répondants. En fait, les panels en ligne importants permettent aux chercheurs d'identifier les membres de sous-groupes relativement rares. En outre, les enquêtes en ligne semblent comporter de bonnes propriétés de mesure; les données de ces enquêtes peuvent être précises, fiables et valides, peut-être même davantage que celles obtenues au moyen des modes plus traditionnels de collecte.

L'effort en vue de rendre plus humaines les enquêtes en ligne peut entraîner une réduction de ces avantages. Dans trois comparaisons sur quatre, Fuchs (2009) a démontré que les femmes étaient significativement plus susceptibles de révéler des renseignements sexuels de nature délicate (par exemple, si elles avaient déjà eu une MTS) à une intervieweuse virtuelle qu'à un intervieweur virtuel; la tendance était moins claire chez les hommes. Il semble probable que plus les intervieweurs virtuels ressemblent à des intervieweurs vivants, plus il est probable que les effets de désirabilité sociale et de l'intervieweur reviennent.

Bibliographie

Bälter, K.A., Bälter, O., Fondell, E. et Lagaross, Y. T. (2005). « Web-based and Mailed Questionnaires: A Comparison of Response Rates and Compliance ». *Epidemiology* 16 : 577 à 579.

- Bason, J. J. (2000). « Comparison of Telephone, Mail, Web, and IVR Surveys of Drugs and Alcohol Use among University of Georgia Students ». Document présenté à la 55^e Conférence annuelle de l'American Association for Public Opinion Research, Portland, Oregon.
- Bates, S. C. et J. M. Cox (2008). « The Impact of Computer versus Paper-Pencil Survey, and Individual versus Group Administration, on Self-Reports of Sensitive Behaviors ». *Computers in Human Behavior* 24 : 903 à 916.
- Chang, L. et J. A. Krosnick (2009). « National Surveys via RDD Telephone Interviewing versus the Internet: Comparing Sample Representativeness and Response Quality ». *Public Opinion Quarterly* 73 : 641 à 678.
- Couper, M.P., F. G. Conrad et R. Tourangeau (2007). « Visual Context Effects in Web Surveys ». *Public Opinion Quarterly* 71 : 91 à 112.
- Couper, M. P., R. Tourangeau et K. Kenyon (2004). « Picture This! Exploring Visual Effects in Web Surveys », *Public Opinion Quarterly* 68 : 255 à 266.
- Denscombe, M. (2006). « Web-Based Questionnaires and the Mode Effect: An Evaluation Based on Completion Rates and Data Contents of Near-Identical Questionnaires Delivered in Different Modes ». *Social Science Computer Review* 24 : 246 à 254.
- Eaton, D. K., N. D. Brener, L. Kann, M. M. Denniston, T. McManus, T. M. Kyle, A. M. Roberts, K. H. Flint et J. G. Ross (2010). « Comparison of Paper-and-Pencil versus Web Administration of the Youth Risk Behavior Survey (YRBS): Risk Behavior Prevalence Estimates ». *Evaluation Review* 34 : 137 à 153.
- Fuchs, M. (2009). « Gender-of-Interviewer Effects in a Video-Enhanced Web Survey: Results from a Randomized Field Experiment ». *Social Psychology* 40 : 37 à 42.
- Kreuter, F., S. Presser et R. Tourangeau (2008). « Social Desirability Bias in CATI, IVR and Web surveys: The Effects of Mode and Question Sensitivity ». *Public Opinion Quarterly* 72 : 847 à 865.
- Link, M. W. et A. H. Mokdad (2005a). « Effects of Survey Mode on Self-Reports of Adult Alcohol Consumption: A Comparison of Mail, Web, and Telephone Approaches ». *Journal of Studies on Alcohol* 66 : 239 à 245.
- Link, M. W. et A. H. Mokdad (2005b). « Alternative modes for health surveillance surveys: An experiment with Web, mail, and telephone ». *Epidemiology* 16 : 701 à 709.
- McCabe, S. E. (2004). « Comparison of Web and Mail Surveys in Collecting Illicit Drug Use Data: A Randomized Experiment ». *Journal of Drug Education* 34 : 61 à 72.
- McCabe, S. E., C. J. Boyd, M. P. Couper, S Crawford et H. D'Arcy (2002). « Mode Effects for Collecting Alcohol and Other Drug Use Data: Web and U.S. Mail ». *Journal of Studies on Alcohol* 63 : 755 à 761.
- McCabe, S. E., M. P. Couper, J. A. Cranford et C. J. Boyd (2006). « Comparison of Web and Mail Surveys for Studying Secondary Consequences Associated with Substance Abuse: Evidence for Minimal Mode Effects ». *Addictive Behaviors* 31 : 162 à 168.
- Schwarz, N., B. Knäuper, H.-J. Hippler, E. Noelle-Neumann et F. Clark (1991). « Rating Scales: Numeric Values May Change the Meaning of Scale Labels ». *Public Opinion Quarterly* 55 : 618 à 630.
- Tourangeau, R., F. G. Conrad et M. P. Couper (2013). *The Science of Web Surveys*. New York: Oxford.
- Tourangeau, R., M. P. Couper et F. G. Conrad (2004). « Spacing, Position and Order: Interpretive Heuristics for Visual Features of Survey Questions ». *Public Opinion Quarterly* 68 : 368 à 393.

Tourangeau, R., M. P. Couper et F. G. Conrad (2007). « Color, labels and interpretive heuristics for response scales ». *Public Opinion Quarterly* 71 : 91 à 112.

Tourangeau, R., M. P. Couper et D. M. Steiger (2003). « Humanizing Self-Administered Surveys: Experiments on Social Presence in Web and IVR Surveys ». *Computers in Human Behavior* 19 : 1 à 24.

Tourangeau, R. et T. Yan (2009). « Sensitive Questions in Surveys. ». *Psychological Bulletin* 133 : 859 à 883.