

ÉVALUATION D'UN PROTOCOLE MULTIMODE AVEC ÉCHANTILLON EMBARQUÉE ET AGRÉGATION DES DONNÉES EN PRÉSENCE D'EFFET DE MESURE

Christophe Barret¹ & Christophe Dzikowski²

- ¹ Centre d'études et de recherche sur les qualifications (Céreq),
Atrium 10.1, 10, Place de la Joliette, BP 21321, 13567 Marseille cedex 02,
barret@cereq.fr
- ² Centre d'études et de recherche sur les qualifications (Céreq),
Atrium 10.1, 10, Place de la Joliette, BP 21321, 13567 Marseille cedex 02,
dzikowski@cereq.fr

Résumé.

Les enquêtes Génération s'appuient, depuis près de 20 ans, sur une collecte monomode téléphonique (Cati). Dans le cadre d'un projet de rénovation du dispositif, le Céreq envisage de basculer sur une enquête multimode internet-téléphone. Après une première expérimentation de collecte monomode internet en 2013, le Céreq a mené une deuxième expérimentation en 2015, basée sur une collecte multimode cette fois-ci. Un double protocole a été mis en place : un échantillon soumis à une collecte multimode « classique » et un échantillon embarqué avec affectation aléatoire du mode de collecte. La collecte multimode classique se déroulait en trois phases : internet seul (avec des relances mail ou téléphone invitant à répondre uniquement sur internet), les deux modes « en compétition » (avec des relances laissant le choix du mode préférentiel) et enfin le téléphone seul (les contacts téléphoniques visaient à répondre directement à l'enquête). L'échantillon embarqué, indépendamment du contact initial (sur le web ou par téléphone), assignait aléatoirement un mode de collecte à l'enquêté. Cela permet en théorie de gommer les effets de sélection lié au mode de collecte et ainsi les écarts observés entre les deux groupes de répondants pourront être interprétés comme des effets de mesure liés au mode de collecte. Cet article a un triple objectif : évaluer l'efficacité des protocoles mis en œuvre, estimer les effets de mesure en utilisant deux méthodes (échantillon embarqué et méthode de matching) et enfin tester différentes méthodes d'agrégation des données issues des deux modes en présence d'effet de mesure.

Mots-clés.

Enquête multimode, internet, téléphone, Cawi, Cati, efficacité du protocole de collecte, échantillon embarqué, effet de sélection, effet de mesure, agrégation des données, ...

1 Introduction

Les enquêtes Génération du Centre d'études et de recherche sur les qualifications (Céreq) permettent d'étudier l'insertion professionnelle des jeunes à l'issue de leur formation initiale. Les jeunes, quel que soit leur niveau de formation, sont interrogés 3 ans, 5 ans et 7 ans après leur première sortie du système éducatif. Ces enquêtes permettent d'alimenter, les acteurs publics ainsi que les chercheurs (en sociologie et en économie) sur le lien entre formation et emploi. Ce dispositif de la statistique publique existe depuis près d'une vingtaine d'années (7 générations suivies) et le Céreq a décidé de s'engager dans un vaste projet de refonte du dispositif. Un des projets est de remplacer le mode de collecte monomode téléphonique par une collecte multimode téléphone-internet. Le Céreq a donc entrepris et poursuit une série d'expérimentations reposant pour tout ou partie sur une collecte internet.

Une première expérimentation de collecte monomode internet (Cawi) a été conduite au printemps 2013 en parallèle de l'enquête principale monomode téléphone (Cati). Les réflexions autour de cette première expérimentation ont été présentées lors du colloque francophone sur les sondages (Dijon, 2015) et des Journées de Méthodologie Statistique (Paris, 2014). Le premier axe de travail concernait l'analyse des abandons en cours de collecte à l'aide de la courbe de « survie » dans le questionnaire. Le second axe de travail a consisté à comparer les estimations entre les deux collectes.

Les structures des réponses des collectes Cati et Cawi obtenues présentaient sur certaines variables des écarts importants, imputables aux effets de mesures et aux effets de sélection. Les effets de mesure caractérisent le fait qu'un même individu réponde différemment selon le mode à la même question. Les effets de sélection proviennent des différences de profil entre les répondants par internet et par téléphone.

Si les effets de sélection posent des difficultés pour le redressement de la non réponse, ils peuvent constituer une aubaine pour les enquêtes car cela permet potentiellement d'en augmenter la couverture. En effet, certains individus ne répondront à l'enquête uniquement s'ils ont la possibilité de répondre sur internet (resp. téléphone). Dans une enquête monomode téléphone (resp. internet), ces individus seraient alors forcément non répondants. Dans une enquête multimode, ces individus auront donc la possibilité de participer à l'enquête (sans que cela nuise à ceux qui répondent indifféremment du mode).

Les effets de mesure sur certaines variables constituent, en revanche, une réelle difficulté à contourner lorsque l'on souhaite produire une estimation en agrégeant les données issues des deux modes de collecte. L'enjeu de la première expérimentation était donc d'évaluer l'ampleur des écarts entre les résultats Cati et Cawi puis d'appliquer des méthodes pour tenter de distinguer ces deux types d'erreurs. La méthode employée était d'isoler les écarts de mesure à l'aide d'une méthode de matching. Si la méthode semble adaptée à ce type de démarche, elle s'est révélée difficile à mettre en œuvre. En effet, d'une part il manquait de variables à disposition susceptibles d'expliquer la sélection (ie. la propension à répondre par internet) et d'autre part, il y avait un problème de robustesse sur l'estimation de l'effet de mesure car les conclusions pouvaient varier selon la spécification du modèle de sélection.

2 Une expérimentation multimode

Une seconde expérimentation, avec une collecte multimode a été réalisée au printemps 2015. Il s'agissait de la deuxième interrogation de la Génération 2010.

Des outils nouveaux ont été mis en œuvre pour essayer de mieux quantifier les effets de mode.

Tout d'abord un échantillon embarqué de petite taille, 1000 individus environ, avec une affectation aléatoire du mode de réponse. Cet échantillon doit permettre d'obtenir deux populations identiques, sans déformation dues aux préférences des individus pour un mode de réponse, répondantes par internet et téléphone et de quantifier, à partir de la comparaison des réponses sur ces deux populations, les effets de mesure dus au mode de réponse. D'autre part, des questions supplémentaires ont été ajoutées dans le questionnaire pour essayer de disposer de variables supplémentaires pour estimer la propension à utiliser internet, et ainsi mieux corriger les phénomènes de sélection liés au mode de réponse.

En parallèle de l'échantillon embarqué avec affectation aléatoire du mode de collecte, une collecte multimode (Cati-Cawi) « classique » a été réalisée sur un échantillon de 4500 individus. L'objectif était d'inciter à répondre en priorité par internet puis d'effectuer un rattrapage par téléphone. Plus concrètement, la collecte s'est déroulée en trois phases de contact successives. La première phase « entièrement internet » consistait à envoyer un mail avis, des relances par mail puis après quelques semaines à entamer des contacts téléphoniques invitant uniquement à répondre au questionnaire web. La seconde phase de relance téléphonique consistait à demander si l'enquêté souhaitait répondre par internet ou par téléphone. S'il préférait internet, un mail automatique était envoyé. S'il préférait répondre par téléphone, l'enquêteur poursuivait directement avec l'enquête. La troisième phase de relance téléphonique consistait à faire directement passer l'enquête par téléphone. Dans cette collecte multimode, l'enquêté avait la possibilité de répondre en plusieurs fois à l'enquête et passer d'un mode à l'autre autant de fois qu'il le souhaitait (le cas le plus fréquent étant un questionnaire démarré sur internet et terminé par téléphone). Un protocole de relance adapté était prévu pour ceux qui avaient interrompu leur questionnaire en cours.

Dans le cadre de ces deux collectes (multimode classique et échantillon embarqué), l'ensemble des contacts se faisaient également de manière multimodal : des lettres-avis (par courrier et/ou par mail), des relances par mail, par téléphone et par SMS.

Pour cette expérimentation, un effort particulier a été consenti pour améliorer l'ergonomie du questionnaire web notamment au niveau du remplissage du calendrier d'activité (qui retrace au mois le mois les différentes situations sur le marché du travail de l'enquêté pendant environ 2 ans) mais également au niveau de l'affichage dynamique (et non plus statique) des consignes.

3 Efficacité du protocole

Une première série de résultats concerneront le protocole de collecte. Quel est l'effet du type de lettre avis, papier ou numérique, sur le taux de réponse. Quel est l'effet de l'envoi de SMS sur le taux de participation ? L'exploitation du fichier historique d'appels (et plus généralement de contacts) permettra à l'aide de modélisation d'établir les effets positifs (ou négatifs) sur la probabilité de répondre à l'enquête.

Un premier modèle pour tester l'effet du SMS de relance, par exemple, semble indiquer un effet bénéfique sur la probabilité de répondre mais paradoxalement cet effet ne se fait pas immédiatement ressentir après l'envoi du SMS (mais seulement plusieurs jours ou semaines plus tard).

Par ailleurs, un des objectifs sera d'évaluer l'efficacité des deux protocoles testés : celui de l'échantillon embarqué où l'on impose à certains enquêtés de changer de mode par rapport au contact initial et le multimode en trois phases (« internet seul », deux modes « en compétition », « téléphone seul »).

Nous tenterons également de mettre en évidence l'effet positif de l'amélioration de l'ergonomie sur les abandons en cours de collecte.

4 Effet de sélection et effet de mesure

La deuxième série de résultats concernera la quantification des effets de mesure et de sélection, et permettra d'avoir du recul sur l'efficacité des outils mobilisés : méthode de matching et échantillon embarqué.

Les résultats issus de la méthode de matching dépendent de la manière dont a été spécifié le modèle. En effet si le modèle est trop spécifié, ou trop précis dans la détermination des effets de sélection aucun appariement ne sera réalisé. Si le modèle ne corrige pas assez les effets de sélection, ces derniers seront interprétés comme des effets de mesures entre les modes de réponses.

L'utilisation de l'échantillon embarqué pour évaluer les effets de mesure semblent plus direct à obtenir et donc plus robuste. La difficulté de cette méthode est ailleurs : en effet le taux de réponse de ce protocole (affectation aléatoire du mode de collecte) relativement faible combiné à un faible échantillon de départ imposera un redressement de la non-réponse efficace et limitera de fait le nombre de variables analysables (variable non filtrée avec peu de modalités).

La comparaison des deux approches sera riche d'enseignements et permettra si les résultats sont proches de potentiellement valider leur usage dans la quantification des effets de mesures.

Une première exploitation de l'échantillon embarqué (EE) compare les réponses des individus affectés aléatoirement au téléphone (EEtel) à celles des individus affectés aléatoirement à internet (EEweb).

Les deux questions suivantes ont été posées à l'enquête principale 2013 (monomode téléphone) et à l'enquête 2015 (affectation aléatoire du mode de collecte : internet ou téléphone). Cela permet donc de comparer les réponses fournies par EEtel et EEweb lorsqu'ils ont tous répondu sur un même mode (téléphone) lors de l'enquête 2013 et lorsqu'ils ont répondu sur un mode imposé (internet ou téléphone) lors de l'enquête 2015.

Comment voyez-vous votre avenir professionnel?					
	Modalités	Répondants EEtel Proportion (en %) (intervalle de confiance à 95%)	Répondants EEweb	valeur test	test égalité des proportions, ($\alpha= 5\%$)
Génération 2010 à 3 ans (tél.)	vous êtes plutôt inquiet?	21,738 (4,465)	25,466 (5,6432)	1,03597	h0 acceptée
	vous êtes plutôt optimiste?	72,341 (4,416)	71,563 (5,778)	-0,2139	h0 acceptée
	Ne sait pas?	5,921 (4,284)	2,971 (1,4652)	-1,30289	h0 acceptée
Génération 2010 à 5 ans (tél. ou web)	vous êtes plutôt inquiet?	11,185 (3,863)	27,499 (5,1089)	5,09402	h0 rejetée
	vous êtes plutôt optimiste?	54,618 (7,678)	43,656 (6,322)	-2,20454	h0 rejetée
	Ne sait pas?	34,197 (7,141)	28,846 (6,2876)	-1,12486	h0 acceptée
<i>Sur échantillon embarqué, données provisoires</i>					

Sur la question « comment voyez-vous votre avenir professionnel? », il n'y a pas de différence significative sur les réponses transmises par EEtel et EEweb lors de l'enquête 2013 (monomode téléphone). Ce résultat permet de se rassurer sur le fait que les deux échantillons permettent de gommer les effets de sélection (puisque sur une enquête monomode, il n'y a pas par définition

d'effet de mesure lié au mode de collecte). Sur l'enquête 2015, en revanche, les répondants internet ont plus souvent tendance à se déclarer « plutôt inquiets » et les répondants téléphone ont plus souvent tendance à se déclarer « plutôt optimistes ». Ce résultat met en évidence un effet de mesure lié au mode de collecte et les écarts bruts entre les deux estimations quantifie l'ampleur de l'effet (+16 points pour « plutôt inquiet » et -11 points pour « plutôt optimiste »). Cet effet de mesure peut s'expliquer par un phénomène de désirabilité sociale. En effet, face à un enquêteur, les enquêtés ont plus souvent tendance à répondre de manière positive ou dans la « norme sociale ». A l'inverse, sur internet, la réponse donnée pourra être plus « sincère » sur une question d'opinion de ce type du fait de l'absence d'interlocuteur.

Envisagez-vous de quitter la région dans les cinq ans à venir?					
	Modalités	Répondants EETel Proportion (en %) (intervalle de confiance à 95%)	Répondants EEweb	valeur test	test égalité des proportions, ($\alpha= 5\%$)
Génération 2010 à 3 ans (tél.)	Oui	33,652 (6,737)	29,375 (6,8737)	-0,88883	h0 acceptée
	Non	66,348 (6,737)	70,625 (6,8737)	0,88883	h0 acceptée
Génération 2010 à 5 ans (tél. ou web)	Oui	22,58 (8,998)	24,853 (9,0361)	0,35639	h0 acceptée
	Non	77,42 (8,998)	75,147 (9,0361)	-0,35639	h0 acceptée
<i>Échantillon embarqué, données provisoires</i>					

Sur ce second exemple, avec une variable plutôt « objective » (moins sujette au phénomène de désirabilité sociale), il n'y a toujours pas d'écart significatif sur la variable de l'enquête 2013 confortant l'idée que les deux échantillons ont bien permis de gommer les effets de sélection. En revanche, cette fois-ci, il semblerait qu'il n'y ait pas d'effet de mesure lié au mode puisque les écarts entre EETel et EEweb sur l'enquête 2015 ne sont pas significatifs.

Ces résultats seront ensuite confrontés à la méthode de matching pour examiner si les résultats sont concordants.

5 Agrégation des données en présence d'effet de mesure

Une fois les effets de mesure mis en évidence et quantifiés. Il est nécessaire d'aborder la délicate question de l'agrégation des données issues des deux modes de collecte. Il existe différentes stratégies d'agrégation des données en présence d'effet de mesure. L'objectif de cette partie sera de mettre en œuvre sur cette expérimentation les différentes méthodes envisagées précédemment (Barret, Dzikowski, 2015), en particulier l'agrégation simple, l'omission d'un mode de collecte et imputation des données omises et enfin la transformation de réponse.

L'objectif sera de déterminer s'il existe une méthode satisfaisante d'agrégation des données en présence d'effet de mesure. Ce point est déterminant dans la décision du Céreq de basculer ses futures enquêtes Génération vers une collecte multimode.

Bibliographie

- [1] Barret C. et Dzikowski C. (2015), La collecte par internet est-elle l'avenir des enquêtes Génération du Céreq ? Comment intégrer les effets de sélection et de mesure liés au mode de collecte ?, *Journées de Méthodologie Statistique (JMS)*, Paris
- [2] Barret C. et Dzikowski C. (2014), Comparaison des modes de collecte par téléphone et par internet dans le cadre de l'enquête Génération du Céreq, *colloque francophone sur les sondages*, Dijon
- [3] de Peretti G., Razafindranovona T. (2014), Les enquêtes multimode : attention aux effets de mode, *Statistique et société, vol2, N°2*
- [4] Anna-Maria Schielicke. (2011). "Don't know" the difference. An experimental comparison between Web and CATI.
- [5] Frippiat, D., & Marquis, N. (2010). Les enquêtes par internet en sciences sociales: un état des lieux. *Population*.
- [6] Krosnick, & Jon, L. C. (2010). Comparing oral interviewing with self-administrated computerized questionnaires an experiment.
- [7] Fougère D (2010), les méthodes économétriques d'évaluation, *RFAS 1-2-2010*
- [8] Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., et al. (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet.
- [9] Annette, J., Caroline, R., & Peter, L. (2008, February). Assessing the Effect of Data Collection Mode on Measurement.