Title: How Do Mode Effects Impact the Measurement of Gender Equality in Household Tasks?

Authors: P. Linh Nguyen, Guillaume Carette, Paul Cochet, Efi Markou, Ruxandra Breda-Popa, Laurent Toulemon

Affiliation: French Institute for Demographic Studies (INED)

Topic description

Gender equality is a key concept in family demography and is of great societal concern in modern times. As its measurement mostly stems from survey data and nowadays surveys are often conducted as mixed-mode surveys, researchers interested in gender equality need to be aware of potential mode effects and their impact on data quality when relying on survey data from different modes. Mixed-mode surveys usually offer one self-administered mode in a cheaper mode, such as the Web, and one interviewer-administered mode (either face-to-face or telephone surveys) to maximise response rates.

Previous research demonstrated mode effects for sensitive, attitudinal questions, such as sexual attitudes, gender discrimination, as well as relationship quality. As much as mode effects vary by question type and content, mode differences can also be differential for men and women with men tending to give more socially desirable answers in interviewer-administered modes such as telephone or face-to-face interviews.

Data

Using the French Generations and Gender Survey (GGS) whose first wave in 2024 was administered as a concurrent mixed-mode survey, we explore mode effects for questions measuring gender equality through the division of various household (HH) and childcare tasks. This measure is the longest item battery within the GGS questionnaire. In the French GGS, the division between cohabiting partners is measured with regard to 8 question items on HH tasks, 5 items on childcare tasks, and 4 items of financial decisions within the HH. For this analysis, we are focusing on the first set of HH tasks and restrict the sample of analysis to heterosexual couples (n=6,563), from which 81% responded on the Web and 19% answered via telephone.

Methods

To examine whether answers differ when an interview is conducted online (Web, self-administered) versus by telephone (CATI, interviewer-administered), a counterfactual framework was chosen: for each respondent, an estimate what their answer would have been under the other mode was constructed via different, complementary propensity-score approaches (matching, stratification, and inverse-probability weighting). Because the same person cannot answer in both modes at once, the propensity score estimates the probability that a person responds online rather than by phone, given observed characteristics (covariates). These characteristics were chosen from both the sample frame (e.g., sociodemographic and HH characteristics) and the questionnaire (e.g., place of birth, educational level, internet access and usage) so that they meaningfully predict mode selection.

Once the propensity scores are modelled, its distribution is compared across modes to identify a region of “common support,” where online and telephone respondents have overlapping scores. Analyses are restricted to this overlap as credible counterfactual comparisons can be made for this restricted group of respondents. These approaches aim to isolate genuine mode effects from differences due to respondent characteristics (structure effects), providing more credible estimates of how Web versus CATI administration influences survey responses.

By conducting analyses using those matching methods between "treated" units (here, web respondents) and "untreated" units (CATI respondents), we obtain an estimate of the Average Treatment Effect on the Treated (ATT), here the effect of answering on the Web instead of the telephone. The results were compared with estimates obtained from a regression weighted by the inverse of the propensity score to assess their robustness.

Our variables of interest are the responses to the eight HH tasks (preparing meals; sweeping the floor; doing the dishes or filling up the dishwasher; grocery shopping; doing the laundry; doing small repairs in and around the house; paying bills and keeping financial records; organising joint social activities) given on a six-points scale (always me; usually me; equally; usually my partner; always my partner; always/usually someone else). We employed multiple correspondence analysis (MCA) to investigate whether the HH tasks division follows an underlying structure with regard to certain similarities in how the eight HH tasks were divided. As per convention, all answer categories with less than 5% of responses were excluded from the MCA. Furthermore, we also set a specific focus on gender-of-interviewer effects analysed through multilevel models, as well as the potential interaction of gender-of-interviewer and gender-of-respondent. We control for the age of both respondents and interviewers through three age groups splitting interviewers in three meaningful tranches and for the survey experience of the interviewer (measured by the year of their first training with the data collection firm).

Results

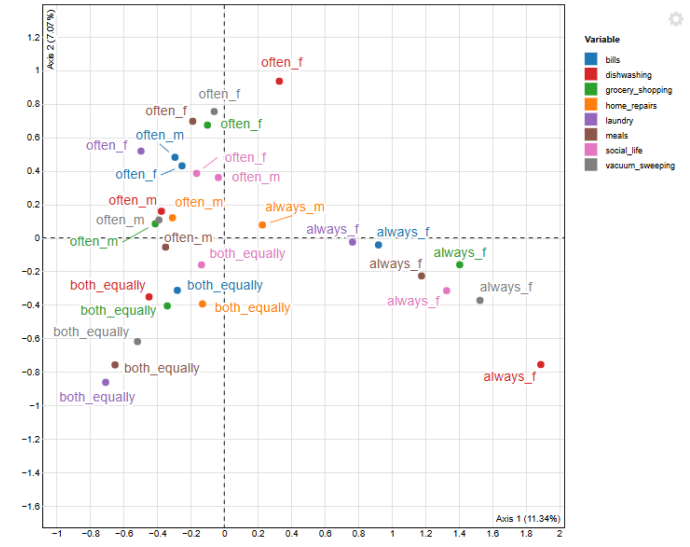

Figure 1. MCA data cloud with answer categories

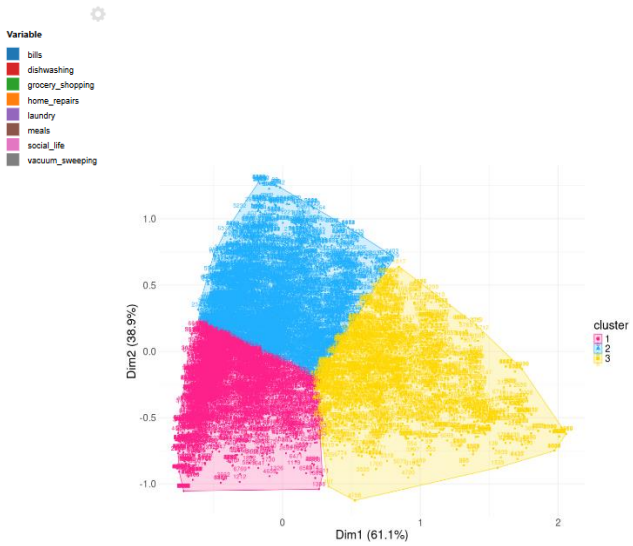

Figure 2. Clustering of data

Figure 1 and 2 show the MCA result rendering three clusters which can be labelled as follows: 1. "Equality within tasks" meaning partners are splitting each task 50/50; 2. "Equality through specialisation" meaning a balanced division of labour by separating who does which task; 3. "Women do most HH tasks" represents a female-dominated division of labour except for doing small repairs. Figure 2 presents the different percentages belonging to a cluster depending on the mode of administration. It seems that respondents declare to share HH tasks rather equally through specialising in certain tasks in Web interviews, while towards an interviewer respondents lean to report that females share the higher burden of domestic work.

Indeed, this difference is confirmed by the ATTs and appears to be similar between male and female respondents (see Table 1). A significant mode effect can be found for females regarding the cluster 2 "Equality through specialisation" (see Table 2). Older respondents over the age of 60 appear to be

the most sensitive to the mode of data collection – a significant mode effect can be found for cluster 1 “Equality within tasks” and 3 “Women do most HH tasks” (Table 3).

Table 1. ATTs across clusters

Clusters	ATT	p-value	Lower CI	Upper CI
1. Equality within tasks	0,005	0,807	-0,034	0,043
2. Equality through specialisation	**-0,047**	0,017	-0,086	-0,008
3. Women do most HH tasks	**0,042**	0,010	0,010	0,074

Table 2. ATTs across gender

	Male respondents				Female respondents			
Clusters	ATT	p-value	Lower CI	Upper CI	ATT	p-value	Lower CI	Upper CI
1. Equality within tasks	-0,005	0,857	-0,063	0,052	0,014	0,598	-0,038	0,066
2. Equality through specialisation	-0,030	0,316	-0,087	0,028	**-0,063**	0,018	-0,115	-0,011
3. Women do most HH tasks	0,035	0,075	-0,004	0,073	0,049	0,059	-0,002	0,099

Table 3. ATTs across age groups

	> 40 years		40-59 years		≥ 60 years	
Clusters	ATT	p.value	ATT	p.value	ATT	p.value
1. Equality within tasks	0,033	0,367	0,035	0,252	**-0,071**	0,046
2. Equality through specialisation	-0,037	0,287	-0,057	0,067	-0,042	0,256
3. Women do most HH tasks	0,004	0,872	0,022	0,384	**0,113**	0,001

Table 4 lists the interaction effects between interviewer and respondent characteristics (divided by gender and age groups). Generally, respondents seem to answer different when faced with female interviewers judging by the smaller ATTs as shown in Table 4. A specific gender effect arises when women are interviewed by a male interviewer as they report an unequal distribution significantly more often – here apparent through the negative ATT in cluster 1 (“Equality within tasks”) and a positive ATT for cluster 3 (“Women do most HH tasks”). Respondents aged 60 and over also appear more sensitive to the interviewer’s age: CATI estimates sometimes are closer to the ones on the Web when the interviewer is aged 40 or younger, and in other cases when the interviewer is a man aged 60 or older.

While the effects observed at the aggregate level are generally also found within the different subpopulations studied (by gender and age), this is not always the case. For CATI questionnaires, interviewer effects can also be observed, which tend to bring the estimates closer to those obtained through the Web.

The variable measuring survey experience of the interviewers does not seem to play a role for interviewer effects in phone interviews.

Table 4. Interactions between interviewer and respondent characteristics

	1. Equality within tasks	2. Equality through specialisation	3. Women do most HH tasks
Male res. x male int.	*Reference category*	*Reference category*	*Reference category*
Male res. x female int.	0,1.3 (0,57)	-0,14 (0,47)	-0,02 (0,94)
Female res. x male int.	**-0,46 (0,05)***	-0,29 (0,2)	**1,03 (0,0004)*****
Female res. x female int.	-0,14 (0,61)	0,14 (0,59)	-0,01 (0,98)
Res. < 40 years * int. < 40 years	*Reference category*	*Reference category*	*Reference category*
Res. < 40 years * 40 ≤ int. < 60 years	-0,12 (0,72)	0,12 (0,7)	-0,12 (0,79)
Res. < 40 years * int. ≥ 60 years	-0,02 (0,97)	-0,27 (0,56)	0,29 (0,65)
40 ≤ res. < 60 years * int. < 40 years	-0,38 (0,32)	0,26 (0,52)	0,21 (0,67)
40 ≤ res. < 60 years * 40 ≤ int. < 60 years	0,26 (0,53)	-0,24 (0,57)	0,06 (0,91)
40 ≤ res. < 60 years * int. ≥ 60 years	-0,14 (0,81)	0,13 (0,83)	0,08 (0,92)
Res. ≥ 60 years * int. < 40 years	**-0,94 (0,04)***	0,06 (0,88)	**1,12 (0,03)***
Res. ≥ 60 years * 40 ≤ int. < 60 years	0,01 (0,99)	-0,1 (0,82)	0,11 (0,84)
Res. ≤ 40 years * int. ≥ 60 years	0,73 (0,24)	-0,21 (0,74)	-0,66 (0,37)
Survey experience	0 (0,88)	-0,01 (0,36)	0,01 (0,42)