

The Impact of Mixed Modes of Survey Administration in the Generations and Gender Survey Round II: Mode Effects and the Role of Mode Assignment

Martijn Schoenmakers^a

a: Netherlands Interdisciplinary Demographic Institute (NIDI-KNAW)

Throughout the social sciences, the use of surveys to gain information on participants is widespread. Different modes of survey administration are available. The Generations and Gender Survey (GGS), a freely available survey tracking topics such as the life course, well-being, and demographic behavior of participants across many countries, provides an example of these different survey modes. In the GGS, questions were traditionally administered to participants in a face-to-face setting (Computer-assisted personal interview: CAPI) or over the phone (Computer-assisted telephone interview: CATI) by an interviewer. With the increasing availability of the internet and the rising costs of more traditional interview methods, questions in the second round of the GGS data collection are increasingly administered to participants via the web (Computer-assisted web interview: CAWI). This shift in the mode of administration may affect both which participants respond (selection effects) and how participants respond (measurement effects). This study will focus primarily on measurement effects. As the answers of participants may be influenced by measurement effects, an understanding of these effects is critical before conducting any comparison involving groups responding to surveys via a different mode.

Previous research has established several ways the mode of administration may affect the way participants respond to questions. For instance, it has been found that the presence of an interviewer may lead respondents to answer in a more socially desirable way (Lee et al., 2019; Mavletova & Couper, 2013). On the other hand, the presence of the interviewer may motivate participants to consider questions more carefully, and the interviewer may offer explanations for confusing questions (Lindhjem & Navrud, 2011). These factors may partly explain the higher response rate that is typically observed for face-to-face surveys (Daikeler et al., 2020). CAPI interviews thus have the potential to increase data quality (motivational and explanation effects) but may also decrease data quality (social desirability effect).

In general, research on measurement effects has largely dealt with more subjective indicators. Indicators traditionally thought to be more objective, such as demographics, may however also be affected by measurement effects. Previous research suggests that while non-sensitive questions usually were not affected by the mode (Christensen et al., 2014), more sensitive demographic questions (e.g., income or education level) can be affected by the mode of survey administration (Lindhjem & Navrud, 2011). Little research has been done on the influence of mode effects on both strongly subjective and relatively objective demographic questions in the GGS-II. One exception is a technical report which examined the effects of mode of survey administration on non-response, straight lining, and rare answers on a set of questions in the GGS-II for France and Uruguay (Toulemon et al., 2024). However, no distinction was made between subjective and objective questions, and only a limited range of questions was considered.

In the GGS, some further research has been conducted on mode effects. For example, a GGS pilot study has found some evidence for the robustness of relationships between subjective and objective variables across face-to-face and web surveys (Piccitto et al., 2022). A key strength of this study was the random assignment of participants to a CAWI or CAPI mode. On the other hand, the sample size for the CAPI group was rather low in this study ($N < 200$), which could have led to decreased power to detect mode differences. A study by Schumann et al. (2023) using the same pilot data found that respondents answer in a less socially desirable way on various indicators of relationship quality when surveys are administered via CAWI rather than CAPI. While both valuable contributions to the literature, it is as of yet unknown if these findings generalize to the GGS-II as implemented in the countries. In particular, the way participants are assigned to CAPI or CAWI administration in practice (usually not via random assignment) may differ between these pilot studies and the full GGS.

The study of measurement effects in the full survey may be particularly interesting for several reasons. First, the full survey is likely to have a far larger sample size than the pilot study, which increases the detectability of any measurement effects. Second, the larger sample size may also enable the use of more advanced statistical techniques to ensure comparability of data across different modes (e.g., measurement invariance testing using MG-SEM: see Hox et al., 2015 for an example of this approach in mixed mode survey data). Third, the results from the full GGS-II naturally are most relevant to researchers concerned about measurement effects, rather than the presence of measurement effects in a pilot data set they are not studying. Finally, the GGS-II not only provides different modes, but also different ways of participants being assigned to the mode (e.g., random assignment, opting in to a mode of preference, being given the choice to switch mode in a reminder if the initial mode was not successful). This adds an extra dimension to the mode effect challenge: does the way people were assigned to a mode further influence outcomes? The aim of this paper is thus twofold. First, we will take stock of and compare the different ways mixed mode surveys were implemented in the GGS-II. Second, we will zoom in on a case study to further illustrate the presence of mode effects in the data.

Data

To prevent confounding potential cultural effects with a measurement effect, the examination of measurement effects is best done by analyzing data gathered via mixed modes within a single country or territory. While data is still being gathered, as of writing several countries and territories have administered the GGS through the use of mixed methods in sufficient sample sizes to be analyzed. Specifically, these are Uruguay (CAPI: $N = 981$, CAWI: $N = 6211$), Taiwan (CAPI: $N = 3061$, CAWI: $N = 4706$), and France (CAPI: $N = 2653$, CAWI: $N = 10814$). In the full paper, the way participants were assigned to different modes will be detailed. In addition, several other datasets may be added to this list as data collection progresses if sufficient sample sizes per mode are reached.

In each country and territory listed, the GGS was administered. The GGS encompasses a large number of questions and scales. Questions range from relatively

objective demographic questions (e.g., “What language do you most frequently speak at home?”) to more subjective questions measuring attitudes or traits (e.g., “How frequently did you experience the following during the previous week: I felt depressed”). The diverse data present here allows us to assess the presence of measurement effects for relatively objective demographic questions (e.g., language spoken at home), sensitive demographic questions (e.g., income and education), and strongly subjective questions (e.g., depression and loneliness).

Methods

To assess the influence of the mode of survey administration on the survey answers, several methods will be utilized. In the case of single indicator questions, we will first test if the mean of the variables differs across modes (t-test). In many cases, it is likely that the demographic properties of the sample (e.g., age, education level, SES) differ between the CAWI and CATI/CAPI groups due to selection effects. To control for these selection effects, we will incorporate these demographic control variables using regression techniques such as ordinal (for ordered outcomes) and linear regression (for continuous outcomes). The difference between the effects with and without controls may also serve as a rudimentary check for the magnitude of the selection effect. In addition to testing mean differences, we will also test if the distribution of answers differs across modes.

For unidimensional scales consisting of three or more indicators, we will apply MG-SEM (including the control variables) to test for configural, metric and scalar measurement invariance across modes and account for measurement error in the variables.

Finally, we will assess if the different mode assignment strategies to countries result in weaker or stronger mode effects. Naturally, this will be done in an exploratory fashion while keeping the limitations of the study in mind. After all, possible observed differences in measurement effect magnitudes may also be due to e.g., the cultural differences inherent to cross-cultural comparisons rather than a difference in the magnitude of the mode effect.

Expected results

First, we expect that the more objective demographic questions will be relatively unaffected by measurement effects (i.e., non-significant differences between modes after controlling for demographic variables).

Second, we expect that sensitive demographic questions (e.g., income and education questions) will be more affected by measurement effects than non-sensitive demographic questions. In other words, we expect significant effects of mode here.

Third, we expect that the subjective sensitive questions (e.g., depression and loneliness questions) will be most affected by measurement effects. We thus again expect significant effects of mode, which are also substantially larger in size than the effect for sensitive demographic questions.

Fourth, we expect that measurement effects will be larger in countries where participants can choose their mode rather than countries where participants are assigned to

modes randomly.

Case Study

To illustrate our approach, we will use data from Taiwan. We will first provide details on how the participants were assigned to different modes here. Second, we will compare objective demographic questions across modes. Third, we will compare demographic questions known to be sensitive across modes. Finally, we will compare subjective questions across modes. Preliminary analyses suggest large differences between CAWI and CAPI administration of the CES-D depression questionnaire in Taiwan, with CAWI participants endorsing the most frequent level of depression symptoms around twice as often as CAPI participants. In addition, the sensitive demographic variables also appear to be affected, since the lowest level of income is reported six times as often in CAWI compared to CAPI.

References

- Christensen, A. I., Ekholm, O., Glümer, C., & Juel, K. (2014). Effect of survey mode on response patterns: Comparison of face-to-face and self-administered modes in health surveys. *European Journal of Public Health, 24*(2), 327–332. <https://doi.org/10.1093/eurpub/ckt067>
- Daikeler, J., Bošnjak, M., & Lozar Manfreda, K. (2020). Web Versus Other Survey Modes: An Updated and Extended Meta-Analysis Comparing Response Rates. *Journal of Survey Statistics and Methodology, 8*(3), 513–539. <https://doi.org/10.1093/jssam/smz008>
- Hox, J. J., De Leeuw, E. D., & Zijlman, E. A. (2015). Measurement equivalence in mixed mode surveys. *Frontiers in psychology, 6*, 87.
- Lee, H., Kim, S., Couper, M., & Woo, Y. (2019). Experimental Comparison of PC Web, Smartphone Web, and Telephone Surveys in the New Technology Era. *Social Science Computer Review, 37*, 234–247. <https://doi.org/10.1177/0894439318756867>
- Lindhjem, H., & Navrud, S. (2011). Using Internet in Stated Preference Surveys: A Review and Comparison of Survey Modes. *International Review of Environmental and Resource Economics, 5*, 309–351. <https://doi.org/10.1561/101.00000045>
- Mavletova, A., & Couper, M. (2013). Sensitive Topics in PC Web and Mobile Web Surveys: Is There a Difference? *Survey Research Methods, 7*, 191–205. <https://doi.org/10.18148/srm/2013.v7i3.5458>
- Piccitto, G., Liefbroer, A. C., & Emery, T. (2022). Does the Survey Mode Affect the Association Between Subjective Well-being and its Determinants? An Experimental Comparison Between Face-to-Face and Web Mode. *Journal of Happiness Studies, 23*(7), 3441–3461. <https://doi.org/10.1007/s10902-022-00553-y>
- Schumann, A., & Lück, D. (2023). Better to ask online when it concerns intimate relationships? Survey mode differences in the assessment of relationship quality. *Demographic Research, 48*, 609-640.
- Toulemon, L., Charrance, G., Cochet, P., Pardo, I., & Rijken, A. (2024). *Technical Paper on Multi-mode GGS Data*. <https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5066fc417&appId=PPGMS>