Advances in Eye-Tracking and Cognitive Interviewing Methodology: Dos, Don'ts, and Decisions

Keywords: cognitive interviewing, eye-tracking, questionnaire testing.

1. INTRODUCTION

Cognitive interviewing is a questionnaire pretesting method by which researchers gain insights into the cognitive processes of how people interpret and respond survey questions. It is mainly used to reveal and fix problems with survey questions and questionnaire structure, with the ultimate purpose of optimising the data collection instrument. The cognitive interview (CI) is a semi-structured qualitative interview, consequently, CI studies are usually done on small, nonprobability samples. The CI simulates a standard field interview in that draft survey questions are administered to a test subject (TS), who responds them as they would normally do. The goal of a CI is to collect data on TSs' response problems and verbal narratives on their question–response process. Cognitive interviewing is more effective in pretesting interviewer-administered questionnaires. The rise of internet-based self-completion (CAWI) prompted enhancing CIs by eye-tracking, which compensates the weaknesses of CI studies on self-completed questionnaires and sheds light on hidden mental processes. However, initial efforts to integrate CIs with eye-tracking (e.g., Neuert & Lenzner, 2015, 2019; Gravem et al., 2018, 2021) show teething troubles reminiscent of outdated CI methods. Therefore, a robust, scientifically reliable framework is needed for eye-tracking enhanced CI studies (ETCIs). The paper outlines how to incorporate eye-tracking into state-of-the-art CI methodology, providing a list of methodology-driven recommendations on essential decisions to aid experts maintain scientific quality and validity.

2. METHODS

Benefits of eye-tracking in cognitive testing of CAWI questionnaires are indisputable. Its teething troubles appear to stem from failing to give heed to a pivotal debate. Kirsten Miller's (Miller et al., 2011) sharp critique of unscientific CI practices yielding impressionistic, anecdotal results brought about a methodological revolution in cognitive interviewing. It culminated in Miller and colleagues' (Miller et al., 2014) interpretive or Theme Coding (Willis, 2015) method: a Grounded Theory-based (Glaser et al.,1967; Charmaz, 2006), rigorous methodology. Setting the standards in many aspects, this rich, descriptive approach opposed simple problem-seeking reparative aims and emphasised the interpretations and the social–cultural context. It goes without saying that introduction of a new data collection instrument should not disregard methodological standards already in place, as the basic principles remain the same. Therefore, the present paper discusses how to address critical decisions of an ETCI study. In the revision of extant ETCI practices, a 'methodology adaptation' method was used, inspired by the theory adaptation framework (Jaakkola, 2020). Problems, shortcomings, and inconsistencies were identified and resolved by applying the principles of state-of-the-art CI methodology.

3. RESULTS

3.1. How to design the sample and recruit participants?

Sample composition is the most salient aspect decision-making. Even a low-budget, small-sample study may yield rich and robust results by careful sample design – an oft-neglected step. Convenience sampling and sample corruption by enrolling colleagues, experts, etc.

is common in ETCI studies (Neuert & Lenzner, 2015, 2016; Chauliac et al., 2016; Gravem et al., 2021). Contrarily, scientific standards demand a purposive sample considering multiple factors (Willis, 2005; Willson et al., 2014; Collins & Gray, 2015; Mújdricza, 2018). Interlocked or parallel quota design is preferred (Collins & Gray, 2015; Mújdricza, 2018). If the questionnaire covers a wide range of topics, we recommend a complex structure: key criteria for interlocked, auxiliary criteria for parallel quotas. To avoid recruitment bias, at least two recruitment channels should be used (Mújdricza, 2018). Creating a reserve sample with the same quota structure is also needed to handle drop-outs. Administering a screening questionnaire with questions on sample and other eligibility criteria is recommended (Collins & Gray, 2015; Mújdricza, 2018).

Concurrent vs. retrospective protocol and (not) using think-aloud

ETCI structure is determined by the relationship between its two major phases. The eye-tracking phase (when the test questionnaire is completed) and the CI phase can be either sequential or simultaneous, not strictly separated in time. This decision also limits the corresponding applicable verbal data collection methods: retrospective or concurrent protocol, respectively. A concurrent probing is much more efficient in recalling cognitive processes; therefore, they are the most often used in cognitive questionnaire tests (Willis, 2015) and usability tests (Geisen, & Bergstrom, 2017). However, some of its attributes raise questions about its applicability in eye-tracking experiments, as the interviewer should be present and active during the completion of the questionnaire, which implies researcher interference, and may significantly impact the eye-movement data (Pernice, & Nielsen, 2009; Neuert & Lenzner, 2015). The think-aloud technique might reduce interviewer-induced bias, as well as provide concurrent data on the TSs' thought processes. Therefore, it may be tempting to ask the TS to voice their thoughts aloud as they fill out the questionnaire (see Nichols et al., 2020). Still, think-aloud is not recommended during the completion phase, and most ETCI studies do not use it. First, many TSs are uncomfortable with thinking aloud, which poses significant risk to data quality. Second, thinking aloud might interfere with the task (Willson et al., 2014; d'Ardenne, 2015). Third, most TSs need to be 'trained' for it (Willis, 2005; d'Ardenne, 2015), which might further affect their natural thought processes. Last but not least, thinking aloud alone in a room is rather unnatural in itself, which compromises the simulation. It is no wonder think-aloud faded from the standard methods of regular CI practice (Willson et al., 2014) as well.

Also related to the considerations above, another question on how to plan a CI study resurfaced with the introduction of eye-tracking. Qualitative research standards (Lincoln & Guba, 1985, Kvale, 1996) disavow the presence of extra observers: it might bias responses or otherwise compromise the interview. This principle is reflected in recent CI standards (OMB, 2016), countering past practice (Collins & Gray, 2015; Willis, 2015). As for recent ETCI studies, various observation practices have been applied. In some cases, the researcher was physically present during completion (Höhne, 2019; Nichols et al., 2020) or observed through a one-way mirror (Bergstrom et al., 2016), while other studies took advantage of the technical capabilities of eye-tracking devices and used a separate observation room (Neuert & Lenzner, 2019). The latter two practices are in line with CI standards, but the former cannot be recommended. TSs should be alone during completion, for quality data can only be collected if the real-life situation is simulated as accurately as possible. Since the physical presence of an observer may affect the TSs' behaviour, it risks inducing bias, and thus compromising the data. Therefore, a two-phase, semi-sequential data collection protocol is recommended: the TS completes the interview alone, real-time observation happening in/from a separate room, which provides the ensuing one-on-one retrospective CI with clues for probing. This way, eye-tracking can compensate for the lack of nonverbal clue detection in a regular concurrent CI setting. Monitoring the TSs'

facial and vocal expressions through a camera is also recommended (see e.g., Neuert & Lenzner, 2019).

3.2. How and what to observe while test subjects complete the questionnaire

Observation should be done employing two additional computer screens: one for the eye-movements and another for the webcam stream (Neuert & Lenzner, 2019). However, not all gazes or fixations bear importance for pretesting purposes. To aid capturing and separating notable events, the observers should have a list of conspicuous eye movements: long fixations, skipping or disregarding a survey component, regressive saccades, etc. Eye movement patterns may differ across TSs due to individual reading skills and habits (Höhne, 2019). It is thus recommended that TSs start the session with reading a simple text, which can serve as a benchmark of their individual reading patterns that the observers can keep in mind during the observation.

If the observation is not assisted by computer reporting, contrarily to the interviewer-only protocol (Neuert & Lenzner, 2019), we suggest that it should be done by at least a team of three: the interviewer and two observers (the team can be smaller if a computer report can be generated on peculiar eye-movements on the spot). A lot can happen, hence, be missed in the seconds of note-taking, so the interviewer's sole task is to watch and narrate the eye-movements without taking their eyes off the screen. The narration of the eye movements should focus on peculiarities that one of the observers can record with predefined signs on a printed questionnaire. Narration allows the interviewer to be fully informed for the ensuing CI and helps etching events in the interviewer's short-term memory, which makes it easier to use the notes and formulate effective probes in the CI. Accurate and prompt recording of peculiar eye-movements can be a rather challenging task in itself, therefore, the task of the other observer is to record the TS' responses. Watching the second screen and note the TS' nonverbal expressions is also the second observer's task. Although seamlessly performing this protocol requires practice, it prevents missing potentially important eye movements.

3.3. Probing: what to ask in the interview

After the TS has completed the questionnaire, the notes should be merged and shortly discussed by the team. Then the interviewer conducts the retrospective, one-on-one CI. In the course of the interview, besides the most common retrospective techniques (scripted and spontaneous probes (Willis, 2005; d'Ardenne, 2015), two additional type of probes appear to be useful. First, it is recommended to ask questions on general (user) experiences at the beginning of the interview, for ignoring a negative experience or disturbing stimulus that may have affected the TS can mislead the analysis. Second, 'semi-spontaneous' [~observational (d'Ardenne, 2015)] probes based on the eye-tracking phase observations are essential for the interview (Neuert & Lenzner, 2019). Theme Coding regards interviewing as the first analytic step (Miller et al., 2014). Considering that similar preliminary eye movement analysis is necessary for efficient probing, the observation phase can be added as a 'zeroth' step to the five-step Theme Coding analysis model. This joint, 'on-the-spot analysis' of eye movements by several observers is a clear advantage of our protocol. The same rules apply to ETCI probes as those of regular CIs: never interpreting the observations, they should be neutral, nondirective, etc. (Willis, 2005, 2015; Willson et al., 2014; d'Ardenne, 2015).

In case the study involves multiple interviewers, interviewing consistency across CIs has to be upheld by adequate interviewer training. Inclusion of eye-tracking in the data collection requires additional preparation and practice regarding the narrated observations

and the probing derived thereof. As a joint exercise, in order to maintain a consistent level in the interviewers' preparation, interviewers (narrators) should interpret at least one eye-tracking footage together. Furthermore, a previously compiled list of phrases, commonly used general probes, and expressions might well be useful during the interview as an aid in formulating semi-spontaneous and spontaneous probes.

Memory-joggers can improve the TS' recollection, mitigating a disadvantage of the retrospective CI protocol. The questions can be shown to the TS in a blank questionnaire [either printed (Neuert & Lenzner, 2015, 2019) or on a display]. Proper application of such tools in a CI as well as directing the TS' attention in the desired way also requires practice. Due to the complex and demanding task of ETCIs, it is essential to do full practice runs prior to the start of the data collection. In theory, showing the TSs cues from their gaze video could also aid probing. However, not only it may distract the TSs but they might also start fabricating explanations (Neuert & Lenzner, 2016), that is, take an 'evaluator' instead of the required 'story-teller' role (Miller et al., 2014).

3.4. How to analyse and interpret the data

Extant ETCI data analysis practices have two major issues. First, they are inclined to conduct quantitative analysis across the whole sample (e.g. Neuert & Lenzner; 2015, 2019), which is problematic on multiple levels. Quantitative results cannot be generalised as they only apply to the sample due to the nonprobability design (Lewis et al., 2014), rendering them effectively useless. CI findings are factual, not statistical (Miller et al., 2014; Mújdricza & Földvári, 2018); therefore, instead of prevalence-based prioritisation, equal importance should be attributed to each observation. Quantitative analysis should be restricted to within-case analyses of deviations from TSs' characteristic reading patterns as it might help spotting less obvious problems and prevent false alarms. Optimally, such individual reports are generated on the spot to assist CI probing. However, this is not always feasible, for the entire questionnaire has to be prepared for the eye-tracking software with its components thoroughly (pre)defined. Second, some ETCI studies use eye movement data to validate and visually underpin CI analysis results, and identify 'main' difficulties (Neuert & Lenzner, 2019). Conversely, the two datasets should be used in a complementary manner: findings, regardless of the method that captured them, should be treated as facts in themselves. Their appearance in both datasets helps to better understand, not validate or prioritise among them.

Theme coding assumed a bottom-up approach to CI data analysis, moving away from the predefined four-stage question–response model (Tourangeau, 1984; Snijkers, 2002; Willis, 2005). The five-step analysis of Theme Coding enables an in-depth data reduction process, moving back and forth between raw textual data, summaries, themes emerging thereof, and conceptual claims (Miller et al., 2014). ETCI data is mainly used in the first (and 'zeroth', observation & interview) and the second (summaries for each question) step. The latter should be divided in 2 phases. The first phase is a 'raw' analysis of eye-tracking data: thick description of observations and preliminary inferences drawn thereof. These are integrated with the CI data in the second phase. Verbatim interview excerpts are summarised so they reflect on and make sense of the findings of the first phase. Phenomena that eye-tracking cannot or did not capture (unique interpretations, problems, etc.) are also important to extract from the textual data. As for problems or other phenomena that are present in only one of the datasets, analysis is usually more informed regarding those identified in the interview excerpts. They have the potential for drawing scientifically sound inferences from them. In contrast, peculiar eye movements alone do not provide insight into their reasons (Chauliac et al., 2016), only allow hypotheses. However, they may shed light on noncognitive mental processes that TSs are unable to reflect on (Neuert

& Lenzner, 2019). For instance, in the 2020 Hungarian Census Test, reverse saccades and long fixations on the reference period of a question was an overarching pattern, though problem was not reported in the CIs despite targeted probing. Our hypothesis was that the numeric format did not align well with the way people recall events of a fixed time period. This observation alone lead to changing the numeric format to a textual reference ('in the last 7 days of April'), which is closer to everyday language use.

4. CONCLUSIONS

Capturing notable eye-movements do not necessarily indicate problems (Neuert & Lenzner, 2019). Interpretation of eye movements may be arbitrary or speculative without a CI and an adequate protocol. The paper presented a 'how to' guide for designing and conducting ETCI studies by a methodology-driven overview of essential decisions. Their discussion also revealed the true potential and limitations of eye-tracking enhanced cognitive interviewing and the way the new instrument impacts data collection and analysis.

REFERENCES

Bergstrom, J. C. R., Erdman, C., & Lakhe, S. (2016). Navigation buttons in web-based surveys: Respondents' preferences revisited in the laboratory. *Survey Practice, 9*(1).

d'Ardenne, J. (2015). Developing interview protocols. In: D. Collins (ed.), Cognitive Interviewing Practice. (p. 101–125) London: Sage.

Collins, D., & Gray, M., (2015). *Sampling and Recruitment*, In: Collins, D. (2015). *Cognitive interviewing practice*. (p. 80–100) Sage.

Chauliac, M., Catrysse, L., Gijbels, D., & Donche, V. (2020). It Is All in the" Surv-Eye": Can Eye Tracking Data Shed Light on the Internal Consistency in Self-Report Questionnaires on Cognitive Processing Strategies?. *Frontline Learning Research*, *8*(3), 26-39.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Sage.

Geisen, E., & Bergstrom, J. R. (2017). *Usability testing for survey research*. Morgan Kaufmann.

Glaser, B. G., Strauss, A. L., & Strutzel, E. (1967). The discovery of grounded theory; strategies for qualitative research. *Nursing research, 17*(4), 364.

Gravem, D. F., Berg, N., Berglund, F., Lund, K., and Roßbach, K. (2018). Test report from Statistics Norway's cognitive and usability testing of questions and questionnaires. Appendix B of WP4 Deliverable 3 of the MIMOD project. Deliverable for Work Package IV of the Cooperation on Multi-Mode Data Collection (MMDC) – Mixed-Mode Designs for Social Surveys – MIMOD. Eurostat Accessed 26. March 2021. https://www.istat.it/en/research-activity/international-research-activity/essnet-and-grants

Gravem, D. F., Meertens, V., Luiten, A., Giesen, D., Berg, N., Bakker, J. and Schouten, B. (2019). Final methodological report presenting results of usability tests on selected ESS surveys and Census. Smartphone fitness of ESS surveys – case studies on the ICT survey and the LFS. Deliverable for Work Package V of the Cooperation on Multi-Mode Data Collection (MMDC) – Mixed-Mode Designs for Social Surveys – MIMOD. Eurostat.

Accessed 26. March 2021. https://www.istat.it/en/research-activity/international-research-activity/essnet-and-grants

Höhne, J. K. (2019). Eye-tracking methodology: exploring the processing of question formats in web surveys. *International Journal of Social Research Methodology*, *22*(2), 199-206.

Jaakkola, E. (2020). Designing conceptual articles: four approaches. *AMS review*, *10*(1-2), 18-26.

Kvale, S. (1996). InterViews: an introduction to qualitive research interviewing: Sage.

Lewis, J., Ritchie, J., Ormston, R. & Morrell, G. (2014). Generalising from Qualitative Research In: Ritchie, J., Lewis, J., Nicholls, C. M., & Ormston, R. (Eds.). (2014). *Qualitative research practice: A guide for social science students and researchers*. (p. 347–366.) Sage.

Lincoln, Y. S. and Guba, E. G. (1985). Naturalistic Inquiry. Beverly Hills, CA: SAGE

Miller, K., Willson, S., Chepp, V. & Ryan, J. M. (2014). Analysis. In: S Miller, K., Chepp, V., Willson, S., & Padilla, J. L. (Eds.). (2014). Cognitive interviewing methodology. (p. 35–50). John Wiley & Sons

Madans, J., Miller, K., Maitland, A., & Willis, G. B. (2011). *Question evaluation methods: contributing to the science of data quality*. John Wiley & Sons.

Miller, K., Chepp, V., Willson, S., & Padilla, J. L. (Eds.). (2014). *Cognitive interviewing methodology*. John Wiley & Sons.

Mújdricza, F. (2018). A kognitív kérdőívtesztelés módszertana: mintaválasztás és toborzás. *Szociológiai Szemle*, *28*(2), 4-27.

Mújdricza, F., & Földvári, M. (2018). A kognitív kérdőívtesztelés módszertana: a kognitív interjúk elemzése. *Statisztikai Szemle*, *96*(6), 545-574.

Neuert, C. E., & Lenzner, T. (2015). Incorporating eye tracking into cognitive interviewing to pretest survey questions. *International Journal of Social Research Methodology*, *19*(5), 501-519.

Neuert, C. E., & Lenzner, T. (2016). A comparison of two cognitive pretesting techniques supported by eye tracking. *Social Science Computer Review*, *34*(5), 582-596.

Neuert, C. E., & Lenzner, T. (2019), Use of Eye Tracking in Cognitive Pretests, Mannheim: GESIS – Leibniz Institute for the Social Sciences (GESIS – Survey Guidelines) Accessed 10. October 2022. https://www.gesis.org/fileadmin/upload/SDMwiki/Neuert_Lenzner_Use_of_eye_tracking_13022019cc.pdf

Nichols, E., Olmsted-Hawala, E., Holland, T., & Riemer, A. A. (2020). Usability Testing Online Questionnaires: Experiences at the US Census Bureau. *Advances in Questionnaire Design, Development, Evaluation and Testing*, 315-348.

Office of Management and Budget (OMB), 2016., Statistical Policy Directive No. 2 Addendum: Standards and Guidelines for Cognitive Interviews. Office of Management and Budget. Accessed 11. August 2021. https://obamawhitehouse.archives.gov/sites/default/files/omb/inforeg/directive2/final_addendum_to_stat_policy_dir_2.pdf

Snijkers, G. J. M. E. (2002). *Cognitive laboratory experiences: On pre-testing computerised questionnaires*. Utrecht University.

Pernice, K., & Nielsen, J. (2009). How to conduct eyetracking studies. *Nielsen Norman Group*, *945397498*. Accessed 10. October 2022. https://media.nngroup.com/media/reports/free/How_to_Conduct_Eyetracking_Studies.pdf

Tourangeau, R. (1984). Cognitive sciences and survey methods. In: Jabine, T. B., Straf, M. L., Tanur, J. M. and Tourangeau, R. (eds.) (1984) *Cognitive aspects of survey methodology: Building a bridge between disciplines*, *15*, (p. 73-100.) Washington: National Academic Press

Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Sage publications.

Willis, G. B. (2015). *Analysis of the cognitive interview in questionnaire design*. Oxford University Press.

Willson, S., Miller, K., Willson. (2014) *Data Collection*. In: S Miller, K., Chepp, V., Willson, S., & Padilla, J. L. (Eds.). (2014). *Cognitive interviewing methodology*. (p. 15–33) John Wiley & Sons.