Katja LOZAR MANFREDA, Valentina HLEBEC, Tina KOGOVŠEK, Bojana LOBE*

COMPARISON OF OFFLINE AND ONLINE FOCUS GROUPS FOR PRETESTING SENSITIVE SURVEY QUESTIONNAIRES**1

Abstract. The study compares traditional synchronous offline focus groups with synchronous online focus groups using chat to pretest survey questions on a sensitive topic. Despite expecting higher data quality in the online focus groups due to their private setting, the comparison of 42 focus groups (21 of each type) revealed minimal differences in data quantity and quality. Moreover, the differences observed were attributed to transcript quality, the moderators' experience, and social skills, participant homogeneity, and familiarity among participants, rather than the focus group setting. This suggests online focus groups can provide a viable alternative or complement to offline focus groups for pretesting sensitive survey questions, especially when faced with cost constraints, epidemic, large geographical distances, or other limits on face-to-face interactions.

Keywords: *offline focus group, online focus group, synchronous chat, data quality, pretesting survey questions.*

^{*} Katja Lozar Manfreda, PhD, Associate Professor, Faculty of Social Sciences, University of Ljubljana, Slovenia, e-mail: katja.lozar@fdv.uni-lj.si; Valentina Hlebec, PhD, Professor, Faculty of Social Sciences, University of Ljubljana, Slovenia; Tina Kogovšek, PhD, Professor, Faculty of Social Sciences, University of Ljubljana, Slovenia; Bojana Lobe, PhD, Associate Professor, Faculty of Social Sciences, University of Ljubljana, Slovenia.

^{**} Research article. DOI: 10.51936/tip.62.3.681

¹ The work for this article was supported by the following Slovenian Research Agency (ARRS) programmes and projects: Social Science Methodology, Statistics and Informatics (P5-0168, 2009–2027); Quality of Life of Social Groups (P5-0200, 2014–2021); Slovenian Public Opinion (PS-0151, 2020–2027); Sexuality in Secondary-school Students in Slovenia: Behaviour, Health and Attitudes (J5-5540, 2013–2016) and Internet Research (P5-0399, 2015–2027).

INTRODUCTION²

Ways of harnessing newly emerging technologies have always been an imperative in social science methodology, with the aim to use them to collect, analyse, interpret and present empirical data, whether by modifying conventional approaches or inventing new ones. In this article, the use of online focus groups for pretesting survey questionnaires is considered. While focus groups (FGs) conducted online are already in use for over two decades (Stewart and Shamdasani 2017), not much research has considered using them to pretest survey questionnaires.

The pretesting of questionnaires involves several steps. These range from early-stage qualitative pretesting in cognitive laboratories applying qualitative techniques, such as expert evaluation, eye-tracking, cognitive interviews, or FGs, to later-stage quantitative pilot studies, and split-ballot experiments in the field (Mohorko and Hlebec 2013; Snijkers 2002). In the last decade, new approaches to pretesting – specifically of survey questionnaires in the online environment - have emerged (e.g., Hlebec and Mohorko 2014; Mohorko and Hlebec 2016). For instance, cognitive interviews that have traditionally been conducted faceto-face can now be implemented remotely, with the interviewer and respondent connected via chat, audio, or video conferencing systems. With web questionnaires, cognitive interviews can be self-administered as respondents type their comments to the questions (e.g., paraphrasing, retrospective thinking, thinkaloud) directly in the online questionnaire. These questionnaires also allow for a more integrated expert review by way of online commenting on the questionnaire, as enabled by advanced web survey software tools (Callegaro et al. 2015, 107). Significant possibilities also exist for split-ballot experiments in the online environment. Finally, online FGs, for which the moderator leads a semi-structured discussion on a questionnaire topic, can be implemented using instant messaging applications (i.e., chat, only allowing the exchange of text), audio or videoconferencing systems, and modern software support (Lobe 2017).

Emphasis in this article is given to FGs as a method of pretesting survey questionnaires. FGs were chosen for this study because the voice of the target respondents themselves (not experts) was of interest, and also since the potential group dynamics of FGs (Stewart and Shamdasani 2014) were expected to provide more valuable results than comments in cognitive interviews or comments made directly in a questionnaire by individual target respondents.

More specifically, our concern in this article lies with the value of online FGs using chat to pretest survey questions compared to traditional offline FGs. For

Note on ethical approval: Although this article is based on research with human participants (in focus groups), this was a non-intervention study and no personal data were collected. Under Slovenian legislation, the approval of an ethical committee is not required outside of the field of medical research, nor are restrictions imposed to prevent research subjects being identified, as in this case. Nevertheless, the students (moderators) who conducted the focus groups were instructed to obtain the informed consent of the participants.

that purpose, we compared a series of traditional synchronous offline FGs and synchronous online FGs using chat that targeted a specific group (i.e., teenagers) to pretest terms used in a particularly sensitive survey questionnaire (i.e., about their sexual habits and health). The results of this comparison are presented below.

OFFLINE VS ONLINE FOCUS GROUPS

Powell and Single (1996, 499) define a focus group as "a group of individuals selected and assembled by researchers to discuss and comment on, from personal experience, the topic that is the subject of the research". When it comes to pretesting a survey questionnaire, the topics discussed in an FG can be terms from the questionnaire, a draft questionnaire, the visual design of the questionnaire etc.

FGs have traditionally been conducted offline whereby a moderator and participants gather in the same room at the same time (synchronous or face-to-face or FTF). Since the advent of the Internet, online FGs have begun to appear. In these groups, the moderator and respondents participate remotely, each by way of their own device connected to the Internet (Lobe et al. 2022). Essentially, online methods facilitate the 'traditional' methods by using infrastructure associated with the Internet (Chen and Hinton 1999, 2). When a FG is conducted online, the aim is to create a computer-mediated "communication event" (Terrance et al. 1993, 53) in an attempt to mimic an FTF interaction format online. The main characteristic distinguishing an online from an offline FG is that the venue being online calls for different skills from both the researcher and the participants (Lobe 2017).

Online FGs can be conducted in various online settings and be classified by the nature of the computer-mediated communication (CMC) as synchronous (e.g., instant messaging (Chen & Neo 2019), audio or video conferencing tools, such as WebEx, ZOOM, MS Teams etc.) or asynchronous (e.g., forums, email) (Jacobson 1999; Mann and Stewart 2000; Nicholas et al. 2010). Communication can additionally be conducted using text-typing (chat), audio or video. However, the discussion in this article is limited to the synchronous chat FGs that were also used in our study. Such an approach resembles 'real-time' data collection (like with an offline FG) since the researcher and the participants are online simultaneously (Lobe 2017). Although this mutual online presence inevitably leads to greater responsiveness and increased interaction, it can sometimes lead to quicker, more superficial answers, and a blurring of the lines between responding to and sending a message (O'Connor and Madge 2003). On the contrary, asynchronous data collection entails a certain time lag between the researcher posting the question and the respondent providing the answer (e.g., via email; Lobe 2017). The time lag between the researcher's and participant's online presence can encourage more exhaustive and reflective answers. In addition, the typing skills of a participant are less of a problem in asynchronous modes.

In the literature (see below), we can find several reviews and empirical studies comparing offline and online FGs. Offline and various types of online FGs are contrasted, generally in terms of cost and time efficiency, implementation issues, and data obtained. Nonetheless, a literature review by Jones et al. (2022) showed there is no "clear consensus as to whether face-to-face or online FGs hold specific advantages in terms of the data produced and the resources required" (Ibid., 1), which calls for further studies on this topic. In this section, we concentrate on the important differences between online and offline FGs while using FGs to pretest survey questionnaires and when synchronous online FGs using chat are involved, and thereby provide the background to our research questions.

Cost and Time Efficiency

When comparing online and offline FGs, the time and cost efficiency of online FGs appears to be the most striking advantage (Namey et al. 2020). Data can be collected considerably quicker and more cheaply (e.g., saving on the time needed to drive to the FG venue, the cost of hiring the venue, avoidance of transcription costs), given that the infrastructure (devices and chat tools) used by the moderator and participants is already widely used by them. Still, if infrastructure (e.g., a virtual platform, webcams) must be purchased, the cost savings might become negligible (Rupert et al. 2017). In addition, if organisational issues are not discussed in detail with participants (e.g., when the FG is organised, how long it will take, how to solve distraction issues) or when administration (preparation) activities (e.g., the uploading of questions onto the chat platform, programming and exchanging electronic consent forms, dealing with unforeseen technologies issues) are not arranged in advance, online FGs with chat can last several hours and might no longer outweigh the transcription savings (Rupert et al. 2017).

Implementation Issues

A noteworthy feature of online data collection is the absence of geographical and temporal limitations. Data can be collected 24 hours a day (Christians and Chen 2004, 19; Joinson 2005, 21). A potentially more diverse population, geographically and otherwise (Joinson 2005, 21; Rupert et al. 2017), can be reached more easily and included than ever before (Coomber 1997, 1). For example, a researcher from Slovenia can set up an online FG with participants based anywhere in the world without having to consider the travel costs, venue, time differences etc. This is especially advantageous when FGs are being used to pretest survey questionnaires with participants from a rare specific population.

The skills needed by moderators and participants are different with online FGs: all are expected to possess at least some level of computer literacy. Further, it is more challenging to moderate an online FG. The online environment, in which participants are physically displaced and more room for disturbances is allowed, brings more challenges to effective moderating. This is especially the

case with asynchronous FGs, as well as when a more structured approach with a carefully moderated discussion and a bigger set of focused questions is required, such as while pretesting survey questionnaires.

The number of participants in FGs is another factor that varies between online and offline modes. In offline FGs, the recommended number of participants with low to medium involvement is 8-10, and 4-6 with high involvement (Morgan and Lobe 2011; Lobe et al. 2022). In online FGs, the number of participants is key to the success of the interaction, notably when the online FG is conducted synchronously in real-time. For example, a discussion in a group with too many participants can move so rapidly that it might skim over complex issues (Mann and Stewart 2000, 113). When the discussion is held in real-time, one can only reply as fast as one can type, which can result in participants who are able to type faster to dominate the discussion. Therefore, not all will initiate, share and participate equally in the discussion. Control over the group's interaction is also considerably more sensitive to the number of participants than with offline FGs (Lobe 2008). Groups made up of 3-5 participants are thus the most appropriate and successful for the online format (Lobe 2008; Lobe and Morgan 2021), albeit the size should also be considered according to the purpose of the FG.

Data Obtained

Greater data accuracy is another advantage of online data collection. If chat (text-typing data collection) is used to communicate, there is no need to transcribe the sessions manually (Christians and Chen 2004, 18; Oringderff 2004, 3) and also no transcription errors. Moreover, data logs gained from FGs can be directly imported into data management packages (e.g., different types of computer-assisted qualitative data analysis software such as CAQDAS). Potentially richer data may also be obtained from the simple "act of having to write", which leads to a more explicit expression of "one's emotions and attitudes" (Joinson 2005, 23).

Nonverbal interaction is often a direct source of data in offline FGs in the form of a probing technique as an alternative to explicitly stated probes (Morgan and Lobe 2011, 219) and by providing valuable contextual information for later data analysis. While in video online FGs, the value of nonverbal interaction may be close to that in offline FGs, this is not the case for online FGs conducted via chat. In fact, CMC interactions in such online FGs have been strongly criticised for lacking visual and social context cues (Kiesler et al. 1984; Kiesler and Sproull 1986; Sproull 1986), ultimately leading to less rich data (e.g., Abrams et al. 2015).

However, the absence of visual interactions means the data are potentially richer due to the openness and intimacy as an outcome of online FG chats being perceived as more private. For example, participants do not hold concerns about their personal appearance and in turn this might encourage certain individuals to participate who normally would not. In addition, when there are no visual cues people can feel in greater control of the way they present themselves, thus adding

to their sociability, friendliness and openness. Individuals freed from being conscious of their appearance are more willing to share their personal feelings and thoughts. Therefore, online FG chat participants might be more open than offline FG participants, and might be more involved in "sharing and comparing" interactions (Walther 1996). The latter aspect is of considerable advantage while using FGs to pretest survey questionnaires on sensitive topics.

SENSITIVE TOPICS IN RESEARCH

In the social as well as natural sciences, topics concerning sensitive issues are quite common, and survey questionnaires are often used to gather related data. In surveys, respondents many times answer such questions with reluctance, partially, or even untruthfully (Tourangeau et al. 2000) and hence unit and item nonresponse and measurement errors are higher. As regards the latter, empirical research shows (e.g., Den Haese and King 2022; King 2022) that respondents tend to underreport socially undesirable behaviours and overreport socially desirable ones (Tourangeau and Yan 2007, 863; Groves et al. 2009). Topics that respondents perceive as sensitive (make them feel 'uneasy') include sexual behaviour, drug use, and other illegal behaviour, unsocial prejudices (like xenophobia, anti-Semitism), income, and even education, profession, spare time, and sports (e.g., Tourangeau and Yan 2007; Krumpal 2023). The methodological literature on sensitive issues is vast and diverse, extending from social desirability concerns to respondents' perceptions of the intrusiveness of particular topics and the possible legal repercussions of disclosing certain information (e.g., admitting to have used illegal drugs; Tourangeau and Yan 2007).

Numerous techniques have been developed over the years to remedy the problems of measuring sensitive issues, such as the self-administration of such questions, applying a randomised response technique, and collecting data in a private setting (Groves et al. 2009). The use of appropriate wording for sensitive concepts (e.g., Uhan and Hafner Fink 2019) can be beneficial. The rise of computer-assisted data collection techniques led to the self-administered completion of web-based questionnaires also being seen as a possible way of reducing problems while measuring sensitive issues (e.g., Booth-Kewley et al. 2007; Tourangeau and Yan 2007; Kreuter et al. 2008; Kays et al. 2012; Krumpal 2023). Several studies have shown that less socially desirable answers are more easily obtained when a questionnaire is completed online than with the paper-and-pencil approach (Callegaro et al. 2015, 24). Yet, one can also find studies (e.g., Dodou and de Winter 2014; Gnambs and Kaspar 2017) that demonstrate the differences between paper-and-pencil and computer-assisted surveys are generally not significant.

When sensitive issues are being studied with qualitative methods, such as in-depth interviews or FGs, the difficulties found in survey research described above seem even more salient, especially because typically there is direct, FTF researcher-participant contact, causing difficulties for both parties. For

instance, researchers encounter several challenges, including the development of rapport, feelings of guilt and vulnerability, leaving the research relationship, and researchers' exhaustion (e.g., Dickson-Swift et al. 2006, Dickson-Swift et al. 2007). As concerns the research participants, especially in a group setting such as an FG, episodes of acting-out or presenting a particular image may occur, alongside participants concealing their vulnerabilities (e.g., Hyde et al. 2005). On the other hand, online FGs, especially chat online FGs, may feel more anonymous, and less awkward (Marley et al. 2023), leading to less embarrassment and social desirability bias. In particular, Samardzic et al. (2024) showed that participants in online FGs – when properly conducted vis-à-vis ethical and security issues and proper facilitations and connections with participants discussing a sensitive topic – were forthcoming and willing to disclose difficult experiences.

RESEARCH QUESTIONS

Our main research question is whether online FGs can complement and sometimes replace traditional offline FGs. The possibility of automated transcripts means that online FGs can save time. Getting people together in a virtual place reduces the time and costs involved with travel and is easier to arrange in terms of their time schedules and other engagements against participation in FG. Online FGs may also be beneficial or even the sole option in a situation like an epidemic or in international research (where the physical distances among research subjects are larger) with limited face-to-face interaction. With respect to sensitive topics, online FGs can potentially yield richer data due to the online interaction being perceived as affording more privacy. All of the reasons described above point to the need to study the value of online FGs with regard to pretesting sensitive survey questionnaires compared to traditional offline FGs.

The above-mentioned difficulties in studying sensitive issues using traditional offline methods led us to compare traditional offline FGs with synchronous online FGs using chat (text-only) as an alternative. We expected that the lack of audio and visual interaction in this type of FGs would give the participants a stronger impression of having greater privacy and a safe space, and would help them open up more about a certain sensitive topic.

Although many scenarios exist for a comparison between the two FG types, the specific features of our study impose certain limitations. In the study, we compared the results of a series of traditional offline and synchronous online chat FGs used to pretest terms from a survey questionnaire on a sensitive issue. More specifically, several moderators each conducted one online and one offline FG and we were able to compare the reports and observations of the moderators themselves as well as the transcripts from these FGs. This led us formulate several questions that can be answered using such a study.

RQ1: Which recruitment mode is used and how successful is it for offline and online focus groups?

One advantage of using an online FG as a pretesting tool may be recruitment. For pretesting, a researcher typically looks for respondents who meet certain criteria. FTF recruitment can be difficult for topics that require specific populations. Another limit is the locality caused by relying on our own social networks and snowball sampling to identify desired respondents so they can convene at the same place to hold an FG. Namely, it is considerably easier to locate participants online who match the highly specialised criteria for purposive sampling (Morgan and Lobe 2011), irrespective of their geographic location. More options are available online for contacting prospective participants in the recruitment process (e.g., not just telephone and email, but also instant messaging, forum private messages, social media sites, etc.). It is also easier to replace drop-outs online (Lobe 2008).

Linked to recruiting suitable participants is the issue of how the group is composed. The aim here is to generate a high-quality group dynamic, which is the main source of data in FGs. With a view to pretesting in online FGs, a researcher has a greater variety of options to assemble an appropriate group since, as already mentioned, it is easier to locate particular types of participants.

In the presented case of pretesting a survey questionnaire, the moderators were free to choose the recruitment mode. Following the literature (Morgan and Lobe 2011), we expected online recruitment to be used more often for the online FGs. We additionally expected that online recruitment would be more successful because a larger number of potential participants was available online.

RQ2: Do sessions of offline and online focus groups vary in their duration? The literature discusses differences in the overall process between the FGs, such as recruitment, setting up the FG, implementing it, and the transcription of data. In this respect, the literature shows this process is usually shorter in synchronous online FGs (Lobe 2008). In our study, while we unfortunately do not have data for the whole FG process, we have data for the duration of the implementation step when the moderator and participants engage in discussion. Here, we expected the online FGs to last longer given the need to constantly type, the more challenging moderation, less control of the process, and greater potential for distractions (Lobe 2008).

RQ3: Which focus group type yields richer data?

The literature (Terrance et al. 1993; Walther 1996; Joinson 2005) suggests that participants of online FGs are more willing to share (in terms of good, quality, unique ideas) than of offline FGs due to the more private setting. This is especially the case (like ours) when a sensitive topic is being discussed by the group. Further, nonverbal interaction is often a direct source of data in offline FGs (Morgan and Lobe 2011, 219) and provides valuable contextual information for later data analysis. In contrast, a CMC interaction in an online FG may lack visual and social context cues (Kiesler et al. 1984; Kiesler and Sproull 1986;

Sproull 1986), what has a significant influence when information is exchanged online. However, more recent work shows that CMC can be depicted as highly socialised, more regulated by norms, and more intimate than FTF interactions precisely because of this lack of social context cues (Joinson 2005, 22). Similarly, Nguyen and Alexander (1996, 104) argue that the lack of visual cues means people can retain greater control of their self-presentation, in turn leading to increased sociability, friendliness and openness. For instance, if participants in an online FG do not think they are being judged by the researcher and their fellow participants according to their physical appearance, they might not worry about this aspect. Walther (1996) suggested that when freed of concerns about our appearance we are able to place stronger focus on the inner self and thus become more willing to share personal feelings and thoughts. Therefore, online chat FG participants might be more open than offline group ones, and might become more involved in 'sharing and comparing' interactions.

RQ4: Which are the main challenges with online moderation compared to offline moderation?

A salient element of a successfully conducted FG is moderation, while the moderating style and its success are crucially important for the data's quality. Offline FGs assembled to pretest a survey questionnaire typically use well-structured and well-tested approaches to moderation (such as providing an open, relaxed, permissive atmosphere; keeping the group on track, and making seamless transitions across topics; encouraging all FGs members to participate, politely closing off the dominant speakers etc.; Groves et al. 2009, 244). Asking questions and managing responses by following the same guidelines should pose no bigger problem in online FGs using chat. Nevertheless, online moderation is in any event more challenging due to the lack of control over the process, the greater potential for distractions, and possibility of blurred interactions (in the sense of who is replying to whom, to which question a reply is given etc.) (Morgan and Lobe 2011). This could, for example, cause bigger difficulties with restoring off-topic discussions or managing overly active exchanges in online chat FGs.

METHODOLOGY OF THE STUDY

We systematically reviewed the results of 21 offline and 21 online FGs conducted by 21 moderators (where each conducted 1 offline and 1 online FG) in the winter of 2014. The aim of these FGs was to pretest a survey questionnaire on the sexual habits and health of teenagers. Although FGs can be used to pretest different aspects of survey questionnaires (e.g., their structure, visual design, instructions), in our case the FGs were chiefly intended to gain insight into the teenagers' understanding of several terms. These varied from less sensitive such as "committed relationship", "to date someone", "to be intimate with someone", "to be a couple", to highly sensitive such as "sex partner", "genital area", "active and passive sexual intercourse", "anal intercourse" etc.

The moderators were international master's students enrolled in a methodology course in which they had been given formal instruction on FG methodology and associated moderating techniques. The course requirements included a task of moderating and transcribing FG sessions. While some students had prior experience in moderation, others were being introduced to this method for the first time during the course, and hence there was a mixed level of expertise among the moderators. Possible implications of this heterogeneity are discussed in the results section. Further, students were also introduced to the aim and topic of the survey questionnaire which they had to pretest. Each master's student had to conduct one online and one offline FG, including participant recruitment and choosing their venue/tool. The offline FGs were implemented in a private (home) or public setting (school, library, bar, meeting room). The online synchronous chat FGs were implemented using Skype, ChatCrypt, Facebook chat, Google talk/hangouts, ChatStep, or E-Chat. Informed consent was obtained from the participants before agreeing to join the study.

Detailed instructions were given to the moderators on the issues to be discussed (questions about specific terms from the survey questionnaire that was being tested); the requested number and structure of participants (4 participants aged 16–18); what to report on the implementation (details of the recruitment process, number and demographics of participants, duration of the FGs, the venue/tools used), and what to report about the results (transcripts, as well as a number of unique ideas, relevant off-topic comments, short statements of agreement, willingness to disclose, nonverbal and paraverbal communication).

Moderators' reports and transcripts from the 42 FGs were coded by 4 coders (authors of this article) who followed a previously agreed coding scheme. The following data were obtained: duration of the FGs, venue/tool for their implementation, number and quality of unique ideas, relevant off-topic comments, short statements of agreement, amount and quality of disclosures, and frequency and quality of nonverbal and paraverbal communication. Some data were missing as an outcome of lower quality transcripts and reports. Apart from the data described above, insights were also obtained from (unstructured and informal) discussions with the moderators. The data were therefore analysed qualitatively and quantitatively, adjusting the analysis to the type of data. Basic descriptive statistics were used for quantitative data whereas a three-step coding process (starting with open coding, followed by axial coding and then selective coding; Straus 1987 in Neuman 2014) was used for qualitative data.

It should be stressed that the focal aim of the presented study was to compare the use of online and offline FGs for pretesting sensitive survey questionnaires and not to present or compare the substantive results of the FGs for them to be used to revise the tested survey questionnaire being tested. Accordingly, in this article we do not report the suggestions on how to revise the survey questionnaire. The fact the transcriptions were made in two languages (Slovenian and English) by the students limited the substantive analysis of the data collected.

Given that several researchers found no significant differences in substantive analysis between online and offline FGs (Nicholas et al. 2010; Synnot et al. 2014; Abrams et al. 2015; Woodyatt et al. 2016; Namey et al. 2020) we also did not focus on this comparison in this study.

RESULTS

Recruitment Process

Moderators were free to choose the recruitment mode. Even though we expected online recruitment to be used more often for the online FGs, this was not the case.

In general, offline recruitment modes were chosen slightly more often, regardless of the FG type (see Table 1). This can be explained by the topic and aim of the FGs: moderators needed to find teenagers to join FGs on a very sensitive topic. We know from discussions with the moderators that most felt it would be easier to approach participants – especially if previously unknown – using FTF interaction as that provides greater legitimacy for the invitation, especially with such a sensitive topic.

Numerous problems with participant recruitment because of the sensitivity of the topic were explicitly reported by the moderators as the biggest reason for refusals, no matter the FG type. The moderators' reports unfortunately do not provide reliable data on the response rates in the recruitment process. However, since regardless of the recruitment mode and FG type all the moderators were able to obtain the requested number and demographic structure of participants, we may conclude that there were no differences in the success of the moderators' recruitment processes.

Table 1: RECRUITMENT METHODS

			FOCUS GROUP TYPE	
			Offline FG	Online FG
Recruitment methods	Offline recruitment methods	face-to-face	4	3
		SMS	2	1
		phone call	3	4
		bulletin board	1	1
		Sum	10	9
	Online recruitment methods	Facebook	2	4
		email	3	1
		online forum	1	3
		Sum	6	8

Source: data from students' reports on focus groups implementation.

Duration of the Focus Groups Sessions

Results concerning the duration of the FGs sessions are in line with our expectations. The online FGs took longer on average (see Table 2, average duration 40.5 min for the offline and 62.7 min for the online FGs). Actually, the duration was longer online for all pairs (all moderators), except one. We had expected a longer duration due to the need to type responses, the more challenging moderation, the lower control over the process, and the greater potential for distractions. Yet, the variability among the moderators (standard deviation of 15.6 min for the offline and 20.5 min for the online FGs) shows the FG duration also depended on the experience and effort of the moderator, not simply on whether the environment was offline or online.

Table 2: DURATION OF THE FGS

		FOCUS GROUP TYPE	
		Offline FG	Online FG
Duration of the FGs (min)	Average duration	40.5	62.7
	Median	40	57
	St. dev.	15.6	20.5
	Min	13	21
	Max	60	90
	N	11	15

Source: data from students' reports on focus groups implementation.

It is worth mentioning that the reports on the offline FGs contained more missing data concerning the duration. Namely, for the online FGs the duration was easier to determine from the full transcript, including the timing of each post, which was recorded automatically by the software used. For the offline FGs, we had to rely on the moderators' reports.

Quality of the Focus Group Data

The results concerning the quality of the data obtained are based on coding the FG transcripts, as well as the moderators' perceptions of data richness. Here, we present results from coding these two data sources in qualitative terms.

Overall, when comparing the offline with the online FGs there seems no definite and clear-cut difference between the quantity and quality of the data produced. Specifically, answers seem generally shorter in the online FGs, although the ideas are not necessarily lower in quality. The shorter answers online seem more to indicate an 'economy' of expression (having to type answers) than the richness of the ideas expressed.

There were almost no off-topic comments, regardless of the FG type. Further, there was more non-verbal and paraverbal communication (or it was at least noted in the transcript as produced automatically) in the online than in the offline FGs. It might be that the respondents in the online groups somewhat compensated for the lack of physical presence by using more elements of electronic non-verbal and paraverbal communication, such as emoticons and other emotional expressions (e.g., written down laughter, words like "uh", "come on" etc.). In contrast, even though nonverbal communication likely occurred in the offline FGs, it was rarely reported in the transcripts.

We found no difference in personal disclosure between the offline and online FGs. On one hand, we expected more personal disclosure due to privacy in the online world potentially being perceived as greater (Joinson 2005). On the other hand, in the online world participants might be more reserved because of the potential for surveillance, data leaks, and hacking (Lobe et al. 2022). Still, the moderators did not report participants' behaviour that would support one expectation or the other. Instead, personal disclosure depended on whether participants knew each other. There is some indication that, at least with such a sensitive topic, respondents who already knew each other might have experienced greater embarrassment and reservation in personal disclosure. The same applies to mixed vs. homogenous groups concerning the sex of the participants. There is some indication of greater embarrassment in mixed groups. The majority of the homogenous groups, irrespective of the FG type, were quite relaxed, with much disclosure, except for two groups where the participants already knew each other. All of the mixed groups, whether the individuals involved already knew each other or did not, showed some embarrassment and less relaxation.

Moderating Style

While previous results were based on the coding of data from the moderators' reports and transcripts, the results presented below are generally based on the moderators' perceptions of the challenges, as written in their reports, and on what the transcripts of the FGs themselves revealed. Here, we present results arising from these two data sources in qualitative terms.

No specific pattern was observed in how the moderators handled the offline and online FGs. In particular, we did not detect differences in group dynamics, restoring off-topic discussions, probing and using follow-up, encouraging participants, minimising distractions, relaxing respondents, managing overly active exchanges etc. In fact, most of these moderation actions almost did not happen, except with some very experienced moderators.

Rather than the setting of the FGs (either offline or online), it was the experience and social skills of the moderators that influenced the quality of the moderation and the success of the FGs. Characteristics such as the moderator's previous (in)experience with the method, social skills, and relaxedness in relation to this sensitive topic seemed to play a crucial role in successfully handling the

FG execution. As concerns all of the previously mentioned moderating activities, the differences among the moderators were actually quite large. For instance, as regards FG moderation, one moderator – in both of the FGs he moderated – was excellent at relaxing the respondents and getting closer to them and motivating them by using examples from his own life. Yet, the same moderator was not so good at bringing off-topic discussions back on track or sometimes even further encouraged them with his own comments. At the other end of the continuum, there were moderators that in both FGs only asked basic questions and almost did not use any kind of moderation techniques, such as motivation, probing, encouragement etc. In most cases, probing was just basic or moderate; good or excellent probing was rare, regardless of the FG type.

It should be noted that while analysing data on the moderating style, we also encountered a problem of the moderators' reports and transcripts of the FGs being of different quality. It could be that in some cases the moderation itself was not as basic as previously mentioned, but that the transcript created later was not as detailed as it should have been. This raises the question of the quality and comparability of transcripts, in turn causing difficulties when evaluating the moderating style.

The variability observed in the moderating style, as well as in the quality of the moderators' reports and transcripts, can be attributed to two core factors: the experience of the students with moderating FGs, and their individual social skills, study motivation, and effort. Our aim was for student moderators to evaluate and compare their experiences in conducting both types (online and offline) of groups. For this reason, we found it more important that they shared comparable levels of preparation and knowledge rather than being highly experienced moderators.

Limitations of the Study

At the time this article was published, the data for this study were over 10 years old. We nonetheless believe the results remain methodologically relevant. Since 2014, and especially after COVID, the use of online FGs has grown considerably. Parallel to this, the number of studies comparing online and offline FGs has also grown. Yet, these studies have yet to provide a clear answer regarding differences in data quality and required resources (Jones et al. 2022). Moreover, we could not find any studies that specifically compare online and offline FGs for the purpose of pretesting survey questionnaires. For example, a literature review on differences between various forms of FGs between 2000 and 2019 included 26 studies (Jones et al. 2022), but none used FGs to pretest a survey questionnaire. From this perspective, our study is unique and useful for researchers intending to collect data using survey instruments.

The study also remains relevant despite the fact that today there are many more tools available for conducting online FGs (e.g., Zoom, Google Meet, MS Teams), which are also more integrated into daily life than they were in 2014.

Namely, fundamental methodological differences between the two types of FGs persist, as outlined in the article (e.g., online FGs offer greater geographic reach, flexibility, and anonymity, while offline FGs provide richer nonverbal communication, higher participant engagement, and less multitasking). Differences in implementation (costs, time) also remain, as well as certain technical challenges (connectivity issues, disruptions, the need for technical skills to be held by moderators and participants), and the need for different skills to moderate online groups (e.g., managing silences, encouraging interaction). Therefore, we believe that the study, even though it is based on data from 2014, still provides a valuable starting point for understanding these challenges today.

DISCUSSION AND CONCLUSION

In this study, we answered research questions related to comparing offline and online FGs based on FGs pretesting a survey questionnaire on a very sensitive topic (sexual habits and health) involving teenagers. The results might accordingly be limited to this specific context.

The answer to our main research question is that synchronous online FGs using chat are at least as valuable as traditional offline FGs in this particular case of studying a sensitive topic using young participants. Namely, when comparing the FGs as to the method used (offline/online), there seems no definite and clear-cut difference in the quantity and quality of the data produced. Some differences, such as the amount of non-verbal and paraverbal communication, are more the result of the quality of transcripts than the FG execution itself. The experience and social skills of the moderators, the homogeneity of the participants, and whether the participants already knew each other are more important to the quality of the results than the offline or online setting.

The fact the overall quality seemed to depend more on the moderators' skills and effort rather than the method (offline/online) applied shows the considerable potential for online FGs to complement or replace traditional FGs for sensitive topics if experienced moderators are involved. We reiterate that our primary goal was to have moderators compare their experiences with each FG type, which was possible notwithstanding their limited experiences and because they were equally skilled. Similar to training interviewers to conduct surveys, great importance and effort should be placed on moderators' training, as well as on instructions that are as clear and detailed as possible for all phases of the implementation of the FGs, perhaps even using exemplary exercise FGs before the 'real' data collection commences.

We conclude by listing the methodological challenges learned during our study. First, the findings are limited to the context of the compared FGs: limitation to a single topic (i.e., a sensitive one) and to a specific population (i.e., young individuals). Pretesting a survey questionnaire on another topic, maybe a less sensitive one, might yield different results. Ideally, a systematic review or even meta-analyses of several more heterogeneous studies would allow for more

generalisable results. Second, experience, training, and clear instructions for moderators are the most important factors that determine the quality of FGs. When large differences exist among moderators, it is difficult to distinguish the impact of different moderators and methods (whether offline or online). Third, there was a lack of certain data on FGs to underpin the analysis and comparison while contrasting the value of online and offline FGs. In our case, for example, we were missing data on the length and costs of the whole process; additional information on the recruitment process (how many participants were contacted, how many refused, how many finally participated); data from transcripts and reports of higher quality. As regards the latter, it could be that in some cases the moderation itself was good, but the transcript done later was not as detailed as it should have been. Some transcripts were highly detailed, noting any pauses, non-verbal and para-language elements, while others showed an (almost) complete lack of such elements. This does not necessarily mean these elements were not present, but more that the moderator failed to note them adequately. This raises the question of the quality and comparability of transcripts, in turn creating difficulties while evaluating moderating styles.

The problem for us as researchers interested in comparing the use of online and offline FGs is that one cannot be always sure what was really going on methodologically in FGs that in substance appear to be 'basic' and to what extent this might have affected the quality and quantity of the collected data. For further research, we suggest giving more detailed instructions to moderators with respect to writing the transcripts for both FG types (apart from text-based, where transcripts are done automatically by the tool used for conducting the FGs) or even ask them to hand in the audio/video tapes rather than the written transcript. This would allow us as researchers to obtain more informative data on the moderating style and its relationship to the quality of the data for both types of FGs. Another interesting line of future research would be to compare video online FG with traditional offline FG, discussing sensitive topics as well as less sensitive ones. This would provide insight into the importance of the impression of having greater privacy in online settings and its impact on data quality.

BIBLIOGRAPHY

- Abrams, Katie. M., Zongyuan Wang, Yoo Jin Song, and Sebastian Galindo-Gonzalez. 2015. "Data Richness Trade-offs between Face-to-Face, Online Audiovisual, and Online Text-Only Focus Groups." *Social Science Computer Review* 33 (1): 80–96. https://doi.org/10.1177/0894439313519733.
- Booth-Kewley, Stephanie, Gerald E. Larson, and Dina K. Miyoshi. 2007. "Social Desirability Effects on Computerized and Paper-and-Pencil Questionnaires." *Computers in Human Behavior* 23: 463–77. https://doi.org/10.1016/j.chb.2004.10.020.
- Callegaro, Mario, Katja Lozar Manfreda, and Vasja Vehovar. 2015. *Web Survey Methodology.* London: Sage Publications. https://study.sagepub.com/sites/default/files/9781473927308_web.pdf.
- Chen, Peter, and S. M. Hinton. 1999. "Realtime Interviewing Using the World Wide Web." *Sociological Research Online* 4: 63–81. https://doi.org/10.5153/sro.308.
- Chen, Julienne, and Pearlyn Neo. 2019. "Texting the Waters: An Assessment of Focus Groups Conducted via the Whatsapp Smartphone Messaging Application." *Methodological Innovations* September-December 2019: 1–10. https://doi.org/10.1177/2059799119884276.
- Christians, Clifford G., and Shing-Ling Sarina Chen. 2004. "Introduction: Technological Environments and Evolution of Social Research Methods." *Online Social Research: Methods, Issues and Ethics*, Mark D. Johns, Shing-Ling Saringa Chen, and G. Jon Hall (eds.), 15–23. New York: Peter Lang.
- Coomber, Ross. 1997. "Using the Internet for Survey Research." *Sociological Research Online* 2: 49–58. https://doi.org/10.5153/sro.73.
- Den Haese, J., and B. M. King. 2022. "Oral-Genital Contact and the Meaning of 'Had Sex': The Role of Social Desirability." *Archives of Sexual Behavior* 51: 1503–08. https://doi.org/10.1007/s10508-021-02220-4.
- Dickson-Swift, Virginia, Erica L. James, Sally Kippen, and Pranee Liamputtong. 2006. "Blurring Boundaries in Qualitative Health Research on Sensitive Topics." *Qualitative Health Research* 16: 853–71. https://doi.org/10.1177/1049732306287526.
- Dickson-Swift, Virginia, Erica L. James, Sandra Kippen, and Pranee Liamputtong. 2007. "Doing Sensitive Research: What Challenges Do Qualitative Research 7: 327–53. https://doi.org/10.1177/1468794107078515.
- Dodou, Dimitra, and J. C. F. de Winter. 2014. "Social Desirability is the Same in Offline, Online, and Paper Surveys: A Meta-Analysis." *Computers in Human Behavior* 36: 487–95. https://doi.org/10.1016/j.chb.2014.04.005.
- Gnambs, Timo, and Kai Kaspar (2017). "Socially Desirable Responding in Web-based Questionnaires: A Meta-analytic Review of the Candor Hypothesis." *Assessment* 24: 746–62. https://doi.org/10.1177/ 1073191115624547.
- Hlebec, Valentina, and Anja Mohorko. 2014. "Effect of a First-Time Interviewer on Cognitive Interview Quality." *Quality & Quantity* 49 (5): 22. https://doi.org/10.1007/s11135-014-0081-0.
- Hyde, Abbey, Etaoine Howlett, Dympna Brady, and Jonathan Drennan. 2005. "The Focus Group Method: Insights from Focus Group Interviews on Sexual Health with Adolescents." Social Science & Medicine 61: 2588–2599. https://doi.org/10.1016/j.socscimed.2005.04.040.
- Groves, Robert M., Floyd J. Fowler Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey Methodology*. London: Wiley.
- Jacobson, David. 1999. "Doing Research in Cyberspace." *Field Methods* 2: 127–45. https://doi.org/10.1177/1525822x9901100204.

- Joinson, Adam N. 2005. "Internet Behaviour and the Design of Virtual Methods." *Virtual Methods: Issues in Social Research on the Internet*, Christine Hine (ed.), 21–34. Oxford: Berg. https://doi.org/10.5040/9781474215930.ch-002.
- Jones, Janet E., Laura L. Jones, Melanie J. Calvert, Sarah L. Damery, and Jonathan M. Mathers. 2022. "A Literature Review of Studies That Have Compared the Use of Face-To-Face and Online Focus Groups." *International Journal of Qualitative Methods* 21: 1–1. https://doi.org/10.1177/16094069221142406.
- Kays, Kristina, Kathleen Gathercoal, and William Buhrow. 2012. "Does Survey Format Influence Self-Disclosure on Sensitive Question Items?" *Computers in Human Behavior* 28: 251–56. https://doi.org/10.1016/j.chb.2011.09.007.
- Kiesler, Sara, Jane Siegal, and Timothy W. McGuire. 1984. "Social Psychological Aspects of Computer Mediated Communication." *American Psychologist* 39: 1123–34. https://doi.org/10.1037/0003-066X.39.10.1123.
- Kiesler, Sara, and Lee S. Sproull. 1986. "Response Effects in the Electronic Survey." *Public Opinion Quarterly* 50: 402–13. https://doi.org/10.1086/268992.
- King, Bruce M. 2022. "The Influence of Social Desirability on Sexual Behavior Surveys: A Review." *Archives of Sexual Behavior* 51: 1495–501. https://doi.org/10.1007/s10508-021-02197-0.
- Kreuter, Frauke, Stanley Presser, and Roger Tourangeau. 2008. "Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity." *Public Opinion Quarterly* 72: 847–65. https://doi.org/10.1093/poq/nfn063.
- Krumpal, Ivar. 2023. "Social Desirability Bias and Context in Sensitive Surveys." *Encyclopedia of Quality of Life and Well-Being Research*, Maggino, Filomena (ed.). Cham: Springer.
- Lobe, Bojana. 2008. *Integration of Online Research Methods*. Ljubljana: University of Ljubljana, Faculty of Social Sciences.
- Lobe, Bojana. 2017. "Best Practices for Synchronous Online Focus Groups." A New Era in Focus Group Research: Challenges, Innovation and Practice, Rosaline S. Barbour and David L. Morgan (eds.), 227–50. UK: Palgrave Macmillan. https://doi.org/10.1057/978-1-137-58614-8 11.
- Lobe, Bojana, and David L. Morgan. 2021. "Assessing the Effectiveness of Video-Based Interviewing: A Systematic Comparison of Video-Conferencing Based Dyadic Interviews and Focus Groups." *International Journal of Social Research Methodology* 24 (3): 301–12. https://doi.org/10.1080/13645579.2020.1785763.
- Lobe, Bojana, David L. Morgan, and Kim Hoffman. 2022. "A Systematic Comparison of In-Person and Video-Based Online Interviewing." *International Journal of Qualitative Methods* 21. https://doi.org/10.1177/16094069221127068.
- Mann, Chris, and Fiona Stewart. 2000. *Internet Communication and Qualitative Research: A Handbook for Researching Online*. London: Sage. https://doi.org/10.4135/9781849209281.
- Marley, Gifty, Rayner Kay Jin Tan, Tong Wang, Chunyan Li, Margaret E. Byrne, Dan Wu, Cheng Wang, Weiming Tang, Rohit Ramaswamy, Danyang Luo, Sean S. Sylvia, and Joseph D. Tucker. 2023. "Online Focus Group Discussions to Engage Stigmatized Populations in Qualitative Health Research: Lessons Learned." International Journal of Qualitative Methods 22: 1–12. https://journals.sagepub.com/doi/pdf/10.1177/16094069231204767.
- Mohorko, Anja, and Valentina Hlebec. 2013. "Razvoj kognitivnih intervjujev kot metode predtestiranja anketnih vprašalnikov." *Teorija in praksa* 50 (1): 62–95.

698

- Mohorko, Anja, and Valentina Hlebec. 2016. "Degree of Cognitive Interviewer Involvement in Questionnaire Pretesting on Trending Survey Modes." *Computers in Human Behavior* 62: 79–89. https://www.sciencedirect.com/science/article/pii/S0747563216301820.
- Morgan, David L., and Bojana Lobe. 2011. "Online Focus Groups." *The Handbook of Emergent Technologies in Social Research*, Sharlene J. Hesse-Biber (ed.), 199–230. New York: Oxford University Press.
- Namey, Emily, Greg Guest, Amy O'Regan, Cristine L. Goodwin, Jamilah Taylor, and Andres Martinez. 2020. "How Does Mode of Qualitative Data Collection Affect Data and Cost? Findings from a Quasi-Experimental Study." *Field Methods* 1: 58–74. https://doi.org/10.1177/1525822X19886839.
- Neuman, W. Lawrence. 2014. *Social Research Methods: Qualitative and Quantitative Approaches*. 7th ed. Harlow: Pearson.
- Nicholas, David B., Lucy Lach, Gillian King, Marjorie Scott, Katherine Boydell, Bonita J. Sawatzky, Joe Reisman, Erika Shippe, and Nancy L. Young. 2010. "Contrasting Internet and Face-to-Face Focus Groups for Children with Chronic Health Conditions: Outcomes and Participant Experiences." *International Journal of Qualitative Methods* 1: 106–21. https://doi.org/10.1177/160940691000900102.
- Nguyen, Dan Thu, and Jon Alexander. 1996. "The Coming of Cyberspacetime and the End of the Polity." *Cultures of Internet: Virtual Spaces, Real Histories, Living Bodies*, Rob Shields (ed.), 99–124. London: Sage.
- O'Connor, Henrietta, and Clare Madge. 2003. "Focus Groups in Cyberspace: Using the Internet for Qualitative Research." *Qualitative Market Research: An International Journal* 2: 133–43. https://doi.org/10.1108/13522750310470190.
- Oringderff, Jennifer. 2004. "My Way: Piloting an Online Focus Group." *International Journal of Qualitative Methods* 3: 69–75. https://doi.org/10.1177/160940690400300305.
- Powell, Richar A., and Helen M. Single. 1996. "Focus Groups." *International Journal of Quality in Health Care* 8 (5): 499–504. https://doi.org/10.1093/intghc/8.5.499.
- Rupert, Douglas J., Jon A. Poehlman, Jennifer J. Hayes, Sarah E. Ray, and Rebecca R. Moultrie. 2017. "Virtual Versus In-Person Focus Groups: Comparison of Costs, Recruitment, and Participant Logistics." *Journal of Medical Internet Research*: 19 (3): e80. https://doi.org/10.2196/jmir.6980.
- Samardzic, Tanja, Cristine Wildman, Paula C. Barata, and Mavis Morton. 2024. "Considerations for Conducting Online Focus Groups on Sensitive Topics." *International Journal of Social Research Methodology* 27 (4): 485–90. https://doi.org/10.1080/13645579.2023.2185985.
- Snijkers, Ger. 2002. *Cognitive Laboratory Experiences: On Pre-testing Computerized Questionnaires and Data Quality.* Utrecht: Utrecht University Repository. https://dspace.library.uu.nl/handle/1874/13401.
- Sproull, Lee S. 1986. "Using Electronic Mail for Data Collection in Organizational Research." *Academy of Management Journal* 29: 159–69. https://doi.org/10.5465/255867.
- Stewart, David W., and Prem N. Shamdasani. 2014. *Focus Groups: Theory and Practice*. 3rd ed. Thousand Oaks, CA: Sage.
- Stewart, David W., and Prem N. Shamdasani. 2017. "Online Focus Groups." *Journal of Advertising* 46 (1): 48–60. https://doi.org/10.1080/00913367.2016.1252288.
- Synnot, Adrienne, Sophie Hill, Michael Summers, and Michael Taylor. 2014. "Comparing Face-to-Face and Online Qualitative Research with People with Multiple Sclerosis." *Qualitative Health Research* 24 (3): 431–38. https://doi.org/10.1177/1049732314523840.

- Terrance, L. Albrecht, Gerianne M. Johnson, and Joseph B. Walther. 1993. "Understanding Communication Process in Focus Groups." *Successful Focus Groups: Advancing the State of Art*, David L. Morgan (ed.), 51–64. Thousand Oaks, CA: Sage Publications. https://dx.doi.org/10.4135/9781483349008.n4.
- Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Tourangeau, Roger, and T. Yan. 2007. "Sensitive Questions in Surveys." *Psychological Bulletin* 133: 859–83. https://doi.org/10.1037/0033-2909.133.5.859.
- Uhan, Samo, and Mitja Hafner Fink. 2019. "Context Effect in Social Surveys: Wording of Sensitive Concepts." *Teorija in praksa* 56 (3): 874–95.
- Walther, Joseph B. 1996. "Computer-Mediated Communication: Impersonal, Interpersonal, and Hyperpersonal Interaction." *Communication Research* 23: 3–43. https://doi.org/10.1177/009365096023001001.
- Woodyatt, Cory R., Catherine A. Finneran, and Rob Stephenson. 2016. "In-Person versus Online Focus Group Discussions." *Qualitative Health Research* 26 (6): 741–49. https://doi.org/10.1177/1049732316631510.

PRIMERJAVA MED TRADICIONALNIMI IN SPLETNIMI FOKUSNIMI SKUPINAMI ZA TESTIRANJE OBČUTLJIVIH ANKETNIH VPRAŠALNIKOV

Povzetek. Študija primerja tradicionalne sinhrone fokusne skupine s sinhronimi spletnimi fokusnimi skupinami (s pisnim klepetom) za testiranje občutljivih anketnih vprašanj. Kljub pričakovani višji kakovosti podatkov v spletnih fokusnih skupinah zaradi bolj zasebne situacije je primerjava 42 fokusnih skupin (po 21 za vsak tip) pokazala minimalne razlike v količini in kakovosti podatkov, zaznane razlike pa so bolj rezultat kakovosti prepisov, izkušenj in socialnih veščin moderatorjev, homogenosti in medsebojnega poznavanja udeležencev kot pa tipa fokusnih skupin. To nakazuje, da so spletne fokusne skupine lahko alternativa ali dopolnilo tradicionalnim fokusnim skupinam za testiranje občutljivih anketnih vprašanj, zlasti v primeru stroškovnih omejitev, epidemij, večjih geografskih razdalj ali drugih omejitev pri osebnih interakcijah.

Ključni pojmi: tradicionalne fokusne skupine, spletne fokusne skupine, sinhroni klepet, kakovost podatkov, testiranje anketnih vprašanj.