

Unveiling the truth: Lessons from a failed list experiment on experiences of wartime sexual violence in Kosovo

Julie Litchfield¹ · Elodie Douarin²

¹University of Sussex Business School

²University College London, School of Slavonic and Eastern European Studies

Wartime sexual violence (WSV) is often stigmatised, and in survey questionnaires, respondents may under-report their victimisation if the question is asked directly. To mitigate this risk, we developed a list experiment allowing victims to report sexual violence experienced during the Kosovo war in a less intrusive, less direct way. While the responses collected through a direct question suggest a prevalence of WSV roughly comparable with previous estimates, our item list revealed an insignificant and negative reporting of WSV. Lower reports of WSV in a non-intrusive question compared to a direct question suggest that the item-list failed. Following a critical discussion of our research design, we explore the underlying patterns in the data and reveal an issue with the data collection itself: when interviewers helped respondents with entering their answers, respondents provided systematically different answers to the count questions. We conclude that variations in the implementation of the interview protocol interfered with our design. Overall, the behaviour of interviewers and interviewees appear to translate complex societal relationship with victimisation. It suggests a possible degree of “fleeing”, a term used to describe a situation where respondents actively avoid being associated with a stigmatised behaviour, but also, and more clearly, an unexpected greater reporting of non-stigmatised victimisation when answers are filled in by an interviewer rather than provided privately. We conclude with suggestions for future work.

Keywords: gender-based violence; wartime sexual violence; conflict-affected settings; list experiment; interviewer bias

1 Introduction

War exacts huge losses, damaging property and infrastructure, disrupting livelihoods and affecting the mental and physical health, and well-being, of survivors. The extant literature suggests that wartime experiences of displacement, injury and death affect a wide range of social, economic and political choices and outcomes in the long-run,

including through the lasting effects of trauma (Bauer et al., 2016, Blattman, 2009, Justino et al., 2013, Douarin et al., 2012). However, some forms of victimisation remain under-researched, particularly wartime sexual violence (WSV). Despite substantial qualitative and anecdotal evidence generated by researchers, journalists and activists during the Kosovo war (e.g. HRW, 2000, KRCT, 2021 and 2023 and di Lellio et al., 2018), reliable quantitative evidence on its prevalence and long-term consequences remains limited. As with other forms of gender-based violence, experiences of WSV are frequently hidden because survivors may fear stigma, shame or social exclusion.

With this in mind, gathering representative and reliable quantitative data on the prevalence and distribution of WSV is valuable but difficult. Direct questioning may lead respondents to conceal experiences they perceive as socially undesirable or stigmatising (Zaal et al., 2026). One ap-

Supplementary Information The online version of this article (<https://doi.org/10.18148/srm/2026.v20i3.8546>) contains supplementary material.

Corresponding author: Julie Litchfield, University of Sussex Business School, Brighton, United Kingdom (Email: j.a.litchfield@sussex.ac.uk)

proach developed to address this problem is the list experiment (or item-count technique), which asks respondents to report the number rather than the type of experiences they have faced from a list of statements. Comparing responses between a treatment group that receives an additional sensitive item, and a control group that does not, allows researchers to estimate the prevalence of the sensitive behaviour while preserving anonymity. List experiments have been used to examine a variety of sensitive topics including racism (Blair and Imai, 2012), corruption (Detkova et al., 2021), tax compliance (Iraman et al., 2022), and attitudes toward intimate partner violence (Asadullah et al., 2021).

More recently, they have been applied in conflict settings to study wartime experiences, including sexual violence (Koos, 2018; Traummüller et al., 2019; González and Traummüller, 2023; Koos and Lindsey, 2022). In three post-conflict contexts—Sri Lanka, Liberia, and the Democratic Republic of Congo—Koos and Traummüller (2022) report substantial evidence that direct questions significantly understate the prevalence of WSV.

Although these studies demonstrate the potential of list experiments for measuring WSV, applications remain relatively scarce. Moreover, existing research has highlighted important methodological challenges concerning list construction, sample size requirements, and the use of direct questions alongside indirect measures (e.g. Blair and Imai, 2012). More recently still, some researchers have called into question the extent to which list experiment could convincingly resolve social desirability bias issues (Blair et al., 2020). For example, while a situation whereby a non-intrusive question elicits less report of a stigmatised experience could suggest an issue with the way the list experiment was designed, a growing set of authors are arguing instead that unanticipated and complex psychological responses to a discussion of the stigmatised item could be at play (e.g. Kotsadam and Løvgren, 2025). Further evidence is therefore required both to assess the effectiveness of the method across different contexts and to better understand the practical challenges of implementing list experiments in studies of wartime sexual violence.

Kosovo is an interesting setting to implement this method. The Kosovo war in 1998–1999 was a short but intense conflict, opposing the Kosovo Liberation Army (KLA) to Serbian forces, following protracted ethnic tensions. The violence of the confrontations prompted mass displacement within Kosovo and an 11-week NATO air campaign in the Spring of 1999. A peace agreement was signed on the 9th of June 1999. Cases of WSV during the 1998–99 war have been well-documented and patterns emerged of trauma made worse by the stigma, shame and taboo of disclosing experiences, often passed down to children (di Lellio et al., 2018; KRCT, 2021; OSCE, 1999; Haxhiaj 2021). Data collected to date on sexual violence

is very patchy but estimates of prevalence nationwide are cited at about 3–4%, 1% of which might have been directed against men (di Lellio et al., 2018). In our informal discussions with local activists and academics, these estimates are however often questioned with most seeing them as underestimating the extent of victimisation but some also expressing concerns over possible over-estimation. In addition, a government program running from 2018–2023 to recognise and provide financial support to victims of WSV has had little take-up, evidencing a persistent stigma over 20 years after the war (Dautry, 2019, Euronews, 2022).

By implementing a list experiment in Kosovo and reflecting on the challenges encountered during fieldwork, this paper contributes to the growing literature on the measurement of wartime sexual violence. Following existing best practice, we administered both indirect and direct measures of WSV (Aronow et al., 2015). Contrary to expectations, the list experiment produced statistically insignificant and negative estimates, while direct questioning yielded a prevalence rate of between 2 and 5%. While the prevalence rate obtained from the direct question is potentially consistent with previous estimates, a lower report from the indirect question was unexpected. A closer examination of the responses revealed that the interview mode was not neutral to the responses collected, and that responses to the item-list were likely affected by social desirability bias. Indeed, while we had planned for the direct and indirect questions to be filled privately by the respondents, in practice they were filled by the interviewers for 43% of our sample. The responses filled by interviewers then show a systematically (and statistically significant) higher count of non-stigmatised forms of victimisation. As we are unable to clarify the reasons behind this selective higher report, we deem our list experiment unsuccessful.

Overall, even if our list experiment did not yield higher reports of WSV victimisation, we feel unable to conclude that the direct responses on WSV provide an unbiased estimate of this specific form of victimisation, because the pattern of responses to our count-item question overall are likely to translate complex social desirability biases. By analysing patterns in responses and drawing on discussions with interviewers, we evidence the sensitivity of the method to small changes in implementation and identify methodological lessons for future applications of list experiments in delicate post-conflict contexts.

2 Research Design

Although the use of WSV during the Kosovo war has been widely commented on (HRW, 2000), the scale of victimisation remains a subject of debate among academics and activists. Experiences of victimisation also continue to be

highly stigmatised and remain taboo for many survivors. This makes it an appropriate case for using a social desirability bias reduction method such as a list experiment. In this section we discuss our efforts to design an item list to measure WSV in Kosovo, building on best practice recommendations.

2.1 List experiment versus direct question

In typical household surveys exploring the degree to which respondents have been affected by war, it is very common to find direct questions about victimisation (Brück et al., 2016). With these questions, respondents are expected to report the negative experiences they have faced during the war. For some more sensitive forms of victimisation, one may however be concerned that some under-reporting can occur in direct questions.

List experiments randomise survey respondents into two groups. The first or control group is presented with a list of items (the “short list”), such as statements about beliefs or behaviours, and asked to report how many, not which, of the items apply to them or they have experienced. The second, or treatment group, is presented with the same list but with an additional item representing the sensitive topic (the “long list”) and asked to say how many of the items they have experienced. Under certain assumptions (Imai, 2011, Blair and Imai, 2012), the difference between the two groups yields an estimate of the prevalence of the sensitive topic item among the respondents. This strategy gives an indication of the extent of under-reporting or bias in estimates based on responses to direct questions. It can also be used to subsequently estimate the effect of the sensitive item on a specific outcome. Indeed, the underlying bias in response captured through a list experiment can be accounted for in follow-up regressions (Tsai, 2019).

In the context of our study, we were interested in providing an updated analysis of the role of victimisation during the Kosovo war on political participation later in life (see Litchfield et al., 2024) by explicitly including WSV as a relevant form of victimisation. Given the stigmatisation of WSV in Kosovo, we opted to include a list experiment that would provide a measure of bias in responses from a direct question and in subsequent regressions, would potentially nuance our findings on the links between WSV and post-conflict political participation by internalising this bias in reporting.

In our study, we included both a direct question about WSV to all respondents and an indirect question as advised in Aronow et al. (2015). We placed the indirect question at the start of the survey, after household demographic information. The direct question was asked at the very end of the survey, as in Igarashi and Nagayoshi (2022), to reduce con-

tamination of responses to the indirect question that might arise if they follow the direct question. Some studies only administer the direct question to the control group (e.g. Agüero and Frisancho, 2022) but in our context where the expected prevalence of wartime sexual violence is low, we opted for making use of the full sample (Blair et al., 2020) to increase the credibility of the responses obtained from the direct question.

2.2 Formulating the direct and indirect questions

Our direct question was formulated as follows:

“Did you or a member of your household experience sexual violence during the war or its immediate aftermath?”

Respondents were invited to respond, yes, no or prefer not to say and could also choose to skip the question (but nobody did).

Our indirect question was presented as follows:

During the conflict, people experienced varied forms of hardships and violence. From the following list, please report how many were experienced by you or someone in your household?

- A Under-nutrition/difficulty in finding food
- B Borrowed money
- C Took step to protect cultural heritage
- D Damage to property and assets
- E [Sexual violence*]

The control group only saw the short list comprising the first four items, while a “treated” group was presented with the long list of five items. While there has been extensive discussion of a possible “optimal length” for an item list in the literature (e.g. Li and Van den Noortgate, 2022) in practice, researchers have often opted for 4 to 5-item list (e.g. see in Koos and Traunmüller, 2022), putting stronger emphasis on carefully selecting the included items.

Our selection of item was designed to reduce the risk of floor or ceiling effects (e.g. Ahlquist, 2018). Specifically, if respondents are highly likely to respond with 0 or the maximum number of items when counting what they have experienced or engaged in, there is a risk that the question will not be perceived as safe and non-intrusive, with instead respondents assuming that their personal experience with the socially undesirable item may then be easily inferred from their response. Here items A, B and D were expected to be common experiences, thus reducing the risk of floor effects (Tsuchiya et al., 2007), while C should have been rare (Freedman, 2024) and thus reducing the risk of ceiling effect. Glynn (2013) also advised including items

not positively correlated with the stigmatised item. Accordingly, items A and B were generalised experiences due to the broad disturbance of war, and thus not necessarily correlated with WSV, while C could be argued to negatively correlate with WSV, because while the risk of WSV increases with being in close proximity with fighters, conservation efforts are more likely carried out by those who are separated from these groups. Item B was chosen as “non-benign”, because reporting financial issues is generally difficult in Kosovo, and Chuang et al. (2021) have reported better performances for list experiments that include a sensitive item which reduces somewhat the salience of the item of interest (here SWV).

Item E, the key item, refers to sexual violence. We deliberately did not narrow this down to only cases of rape and instead adopted the broader definition of sexual violence as rape, sexual assault, sexual slavery, forced prostitution, forced sterilization, forced abortion, and forced pregnancy (UNSG 2019). Note that the survey questionnaire did not explicitly set out this definition although the definition was discussed with enumerators during training, to allow them to discuss it with respondents consistently as needed.

The decision to ask respondents to report an experience with WSV in the household was informed by discussions with researchers familiar with current norms in Kosovo, who pointed at specific ethical concerns: there remains a significant level of stigma associated with WSV, despite public campaigns and the establishment of the government commission to recognise victims of sexual violence. Directing the focus away from the individual respondent to the slightly more anonymous “you or someone in your household” protected the well-being of our respondents. Broadening the question in this way allowed to balance out giving space for victims to have their experience recognised while reducing social stress and risk of re-traumatisation when faced with the direct question. This decision was clearly explained in our ethical approval application. While we recognise that a focus on experience within the household can constitute an increase in the cognitive load associated with answering the indirect question, our focus was on formulating the direct question sensibly and adopting a consistent approach in the direct and indirect questions. The reference to household experience is consistent with the approach adopted in direct questions relating to stigmatised behaviours to facilitate disclosure (see for example the body of work on engagement with bribery and corruption discussed in Amini and Douarin, 2020) and is consistent with the approach adopted (successfully) by Koos and Lindsey (2022) in a list experiment on WSV.

2.3 Interview format

We planned the interview format to provide privacy and protection for the participants. Participants were not to be asked to give a verbal response to any sensitive question, whether direct or indirect, and instead we planned to use self-completion of the questions on all sensitive wartime experiences. Specifically, the interviews were conducted in person, and responses were entered directly on a tablet. Interviewers were instructed to pass the tablet to the participant for them to read the question for themselves and to enter their response for all sensitive questions. This avoids the need for the interviewer to read aloud the question, gives privacy to the participant about their response, and reduces interviewer bias. In other studies, such as Koos and Traummüller (2022), researchers have asked enumerators to turn away from respondents while reading the question to give this privacy. We felt this risked the interview being overheard by others, as interviews were expected to be conducted at the front door of the respondent’s house rather than in a private space.

Training of the enumerators was led by the UBO Research Director and the authors of this paper and was designed to ensure that the interview team understood the sensitive nature of the research. Enumerators were trained to pause during the interviews after any of the sensitive questions, and should the respondent express discomfort or appear distressed. Enumerators were trained on using the item-count question through role-playing, first with other enumerators and then, in a pilot conducted in Pristina before the survey began. This allowed them to gain insights into how to explain the questions and support respondents without being intrusive.

All participants were provided with an information sheet which advised participants that some of the questions related to war-time experiences of injury, death, sexual violence and displacement. Although this potentially risked priming respondents and may have led to a degree of selection bias (Stancheva, 2023), we felt for ethical reasons that respondents needed to be informed about the survey topic and allowed to refuse to take part on this basis.¹ Participation was voluntary; contact details of UBO and the Principal Investigator were provided, and participants were informed that they could withdraw consent up to two weeks after the interview. At the end of each interview, the enumerator provided the participant with a supplementary leaflet containing sources of advice and support.

¹ Refusal to be interviewed was small: a total of 1836 people were invited to participate, of which 261 refused, of which almost half cited lack of time as the reason for refusing, and only 6% saying that they were concerned about data privacy and protection.

2.4 Sampling and statistical power

Regarding sample size, list experiments are typically demanding if the objective is to get an accurate estimate the prevalence of a stigmatised experience which is believed to be rare (Blair et al., 2020). As previously stated, the best current estimates place WSV as having affected about 3-4% of the Kosovar population (di Lellio et al., 2018). While this represents a shocking level of sexual violence, a prevalence of 3% means that only about 30 respondents out of 1000 will respond yes to a direct question on WSV, if they answer truthfully. A much larger sample will be needed to accurately assess the existence of a bias in responses via an item-count question (e.g. see discussion in Blair et al., 2020).

Our budget allowed for a sample of approximately 1500 individuals to be surveyed in November 2023. This sample is not very large, although of similar order of magnitude to those used in other post-conflict contexts and focusing on experiences carrying stigma such as Koos and Lindsey (2022) and Traunmüller et al. (2019) and we used several strategies to increase the likelihood that we would encounter respondents who had been victimised in some ways during the conflict. Rather than drawing a nationally representative sample over the entire country, which risked resulting in low rates of WSV, we selected a set of municipalities most affected in terms of damages suffered during the conflict and oversampled specific sub-groups.² Although a number of cases of WSV have been documented well enough to give an approximate guide to its spatial distribution, incomplete reporting and documenting of cases suggested that a better sampling frame would rely on the wider set of wartime experiences. There are several studies that identify casualties in numbers and rates, and damage to housing from bombing and mines (EC 1999, and sources in Trako, 2018) and di Lellio et al. (2018) document locations of over 200 survivors of WSV. This targeting of municipalities also leads us to focus on regions with large ethnic-Albanian majorities: given recent instability in Serbian enclaves linked to the Russian war in Ukraine, this provided a safer context for enumerators.

Our final sample of 1574 respondents thus consisted of two groups. The first 1069 respondents were from households (one respondent per household) randomly selected to be representative of the adult population in terms of

age, gender and rural/urban residence according to the 2014 Kosovo census³, but exclusively in municipalities with extensive war damages. Each interview was conducted in person, and each respondent was selected through a random walk within primary sampling units in each selected municipality. We refer to this as our main sample. We added to this a second sample of 505 respondents representing an over-sampling from the population widely regarded as having been most at risk of WSV, namely adult women in rural areas (di Lellio et al., 2018) within the same municipalities, who at the time of the survey would be aged over 40. We refer to this as our supplementary sample. This approach was chosen to allow us to make inferences for the population using sampling weights derived from the 2014 census data, while also increasing the chances to have a non-trivial number of WSV cases reported. We note that in this article we focus on interpreting the patterns of responses within our sample rather than on extrapolating findings to the wider population. Unless otherwise stated, our results section thus presents raw data and unweighted analyses.

3 Results

3.1 Prevalence of WSV

The direct question we included in the survey did receive positive answers, showing that even a simple direct question can be helpful in measuring WSV experiences. Specifically, out of 1573 valid responses to whether a member of the household has experienced WSV, we received 35 positive answers (2%), 25 prefer not to say (2%), while the remaining respondents responded negatively. Focusing on women only in the representative sample (518 valid responses), 3% of them reported an experience of WSV, or 5% if we also include those who said they would rather not respond. In our supplementary sample of rural women (505 valid responses), 3% responded yes, and 4% of them responded either yes or prefer not to say.

In the immediate aftermath of the conflict, di Lellio et al. (2018) had reported an estimated 3-4% prevalence rate of WSV across the population. This estimate is contested because it is based on surveys conducted in displacement camps during and shortly after the war, it thus uses data from a restricted section of the population, extrapolated to the overall population. It is still the most reliable and reported estimate. At first glance, our direct question seems to have elicited reasonable responses since the derived raw

² These are Deçan, Gjakovë, Glogoc (now Drenas), Istog, Klinë, Mitrovicë South, Pejë, Podujevë, Prishtinë, Prizren, Skënderaj, Suharekë, Viti, Vushtri, and Mitrovicë North. Mitrovicë is one municipality, however it is divided in two different areas whereby the North of Mitrovicë is more populated by ethnically Serbian citizens, and the South of Mitrovicë is more populated by ethnically Kosovar Albanian citizens. Note that some municipalities had different names at the time of the conflict.

³ While the 2014 Census is dated, the more recent 2022 census was still unavailable when we collected our data. We thus made use of the most up-to-date data available at the time of the data collection.

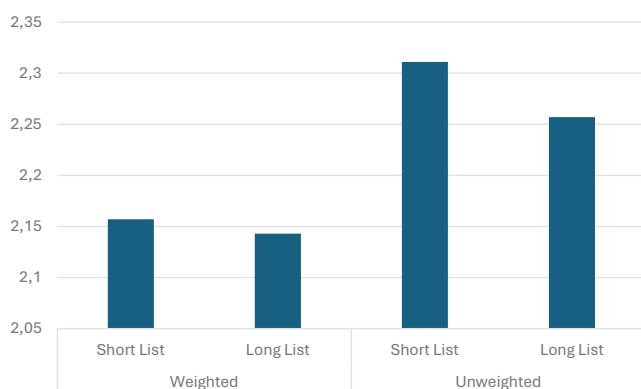


Fig. 1

Mean item counts for short and long list respectively for combined sample of $n = 1573$; weighted and unweighted.

prevalence aligns quite well with this reference estimate. However, by design we focused our efforts on municipalities with more war exposure, and we oversampled older women from rural areas, we were thus expecting to detect a higher rate in the raw data. Our question also focused on experience within the household rather than personal experience only, which could have led to more frequent reporting. At the same time, over 20 years after the end of the conflict, anecdotal evidence strongly suggest that a direct question could only lead to low reporting, due to continued social stigma.

Turning to the indirect question, we start by inspecting the descriptive statistics. Fig. 1 reports raw averages for the item count question. Respondents who were presented our short list of 4 items reported having experienced an average of 2.31 items during the war, while those presented with the longer list of 5 items (i.e. the same list with WSV as a fifth item), only reported having experienced 2.25 items on average. Weighted averages (making use of sampling weights based on the 2014 Census) indicate an average of 2.16 experiences from the short list and 2.14 from the long list. The fact that we find that the mean item count for the long list is lower than for the short list goes against expectations, as at least those respondents who did report WSV in the direct question should also have reported it in the indirect question, leading to an expected higher average report on the long list.⁴

This observation is confirmed by econometric analyses. The estimates produced using the *kict* Stata package designed by Tsai (2019) are statistically insignificant and negative for the full sample. The same holds for the representative sample, the men only in the representative sample

and the women only in the representative sample. For our supplementary sample of rural women aged 40 or over the estimate is positively signed but still insignificant. These estimates are essentially obtained by taking the (conditional) difference between the responses obtained from the long and the short version of our item lists. Thus, a negative estimate suggests that counts derived from the long list have been smaller than the counts derived from the shorter list, a counter-intuitive results given the random allocation of the long-list treatment, and the fact that at least some respondents have been willing to report WSV in our direct question. Statistic descriptives and econometric analyses thus converge in suggesting that our list experiment has not detected any incidence of WSV, but also that something might have gone wrong.

3.2 Validating our design

We explore the three key assumptions about the design and implementation of the experiment as detailed in Tsai (2019). The first is that the treatment, i.e. being asked to count war time experiences using the long list of five items, is randomly assigned to respondents. We tested this across age, municipality and ethnicity, and for the main sample only, gender, using balance tests and found no statistical differences between control and treatment groups, confirming that this assumption holds. This is further illustrated in Figures A1 and A2 in the Supplementary Information, which show the raw differences for these key characteristics for our treatment and control groups.

The second assumption is that respondents answer truthfully. There is no formal test of this but the design of the interviews, with interviewers instructed to hand over the tablet to respondents, should have encouraged truthful reporting, by reducing concerns over stigma. Respondents were also given the option to skip any question to reduce the likelihood of untruthful response, but none of them actually skipped the questions on WSV (direct or indirect). The fact that our direct question elicits prevalence rates that are in line with previous estimates should reassure us that we are getting credible responses elsewhere in the survey. In addition, as our experiment contains both an indirect and a direct question, we have an opportunity to explore the consistency between responses to these two questions.

We have already noted that the placement of our questions means we can discount the possibility that the experiment has been contaminated by priming effects. Then, selecting only those respondents in the treatment group, we compared the mean item count across responses to the direct question. We would expect that those who answer positively to the direct question on WSV should have higher item counts than those who do not, and we do indeed find

⁴ Note that the differences between the short and long list are not statistically significant, based on t-tests.

this to be the case. Of the 765 respondents in the treatment group, those who reported that someone in their household had experienced WSV during the war (responding strictly yes to our direct question) had a mean item count of 3 (out of 5), statistically greater than the mean item count of 2.30 of those who reported no experience of WSV in the direct question.⁵ This does support the design of our item count list itself, as it shows that the respondents who are willing to report WSV in a direct question, also report a higher item count on the long-list. But, the systematically negative estimates, even if not significant, for the indirect question do suggest that answers may have been biased in some ways and calls for further investigations.

The third assumption is that there are no design effects, namely that the addition of the sensitive item does not influence the probabilities of the other items being selected. First, we attempted to minimise design effects by including the sensitive item as the last one on the list. This approach is suggested in Detkova et al. (2021) and assumes that if respondents count the item that apply to them as they go along, the last item will not distort their answers to earlier items. Second, we also test this formally using the *kict* package in Stata 18 again (Tsai, 2019). We find no evidence of design effects in either the combined or the main sample, but we are unable to test this for the smaller supplementary sample. We note however, that this test is not perfect and has been shown to over reject design effects, especially in small samples (Ahlquist, 2018).

3.3 Responses and interview mode

To go further, we looked at the survey responses again. Our fieldwork plan was that interviewers would hand over their tablet to the respondent, providing a degree of privacy, when answering sensitive questions. At training, some enumerators had expressed a concern that some respondents might not be able to read or may be unfamiliar with a tablet, in which case assistance could be offered by the interviewer or by a family member. If this option was chosen, it would be logged. While this was expected to be a remedial solution in a minimum number of cases, we found that only 51% of the responses to the direct and indirect questions were in fact self-completed (with near perfect correlation on the option chosen across both questions), while 43% of the responses were completed with assistance from the interviewer, and assistance from a family member was reported in the remaining cases (6%).

⁵ Including also the respondents who selected “prefer not to say” as having potentially experienced WSV, the item counts are closer together, at 2.69 and 2.23, but this difference is also statistically significant.

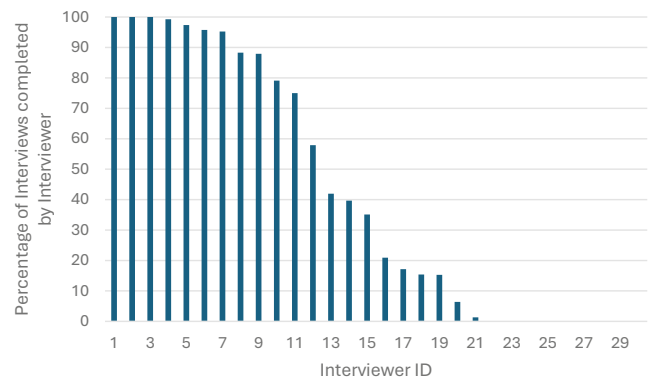


Fig. 2

Percentage of Interviews Administered by the Interviewer. Interviewer ID numbers have been anonymised and randomly reassigned to protect interviewer identity.

Furthermore, we note that some interviewers were much more likely than others to not pass the tablet over to the respondent. Fig. 2 shows the proportion of interviews where the response mode was interviewer assisted, at the level of each interviewer. Eleven of the thirty enumerators completed over 75% of the responses themselves, and seven of these completed 95% or more of the responses themselves.

Since the identity of the interviewer seems to loom large on the decision to get helped to fill the sensitive questions, for the rest of this section we will compare the responses of those who were helped by the interviewers versus the rest, namely self-completion and helped by a family member.

Following the same logic as in section 3.1. above, we present some descriptive statistics first before moving on to

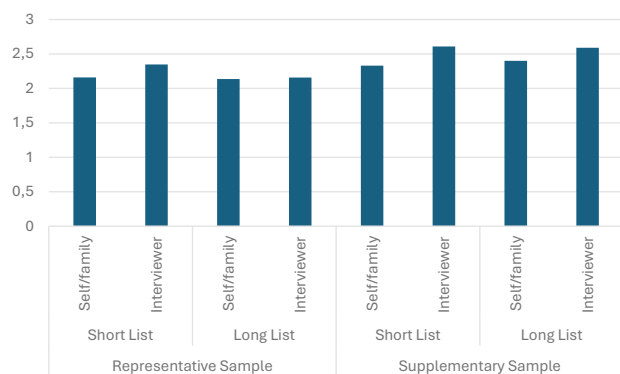


Fig. 3

Mean item count (unweighted) for short and long list respectively for Representative and Supplementary Samples, by mode of interview, i.e. Self-completed or assisted by a family member vs assisted by the interviewer.

our econometric analyses in Sect. 3.5. Fig. 3 below shows the average unweighted item counts for the treatment and control groups disaggregated further by the sample type and by interview response mode. A pattern emerges whereby the reported item count on the short or long list is greater when the response was entered by the interviewers themselves for nearly every sub-sample. While the differences are small, these patterns seem to suggest that the interview mode interfered with how respondents answered the indirect question.

3.4 Focus group with the enumerators

After fieldwork had been completed and the data had been fully reviewed, we organised a focus group meeting with some of the enumerators. To avoid any sense of blame, we asked the enumerators to reflect on reasons why either they, a colleague or a hypothetical enumerator might not have handed over the tablet and what might be done for future surveys to avoid this. Several themes emerged from these discussions.

Firstly, and as anticipated, technology barriers were mentioned by all enumerators, especially for older and less literate respondents, who they expected would struggle to read or enter a response into a tablet. Older respondents, respondents with lower levels of education or respondents with less digital literacy, something that is typically more likely for older generations and in rural areas in Kosovo, may have thus been more inclined to ask the interviewer for help with filling the survey.

Secondly, the enumerators provided additional insights that are informative. All felt that part of the explanation for the tablet not being handed over is that respondents felt very comfortable and trusting of the enumerators. Having explained the purpose of the survey and shared the Participant Information Sheet which clarified that sensitive questions might be asked, interviewers believed that the respondent trusted them enough to suggest they just continue holding the tablet. One enumerator explained that insisting they hand over the tablet after it had been turned down felt rude and put pressure on the respondent which might then affect their engagement with remaining questions. There is no evidence in the data collected that non-responses were more frequent when answers were entered by the respondents themselves. And it is also certainly the case that the team of enumerators are well trained and experienced in running surveys, including on sensitive topics. It is thus possible that the selection of the interview mode was as natural and neutral as some of the interviewers seem to believe.

Thirdly, the enumerators said that some respondents thought it would be quicker to not have the tablet handed

over. This is an important consideration as we did not offer any compensation to respondents for their time. We do find that interviews that were self-completed took slightly longer (17.9 min compared to 16.6 min) and although this is a small difference, it is statistically significant. Thus, finding ways to speed up completion might be useful, and in hindsight, we could have reflected more on providing incentives for enumerators to comply with interview protocols, or to respondents to spend time on our survey.

Fourthly, another insight from the discussion with enumerators was that many respondents, particularly those in areas where cases of WSV have been documented appeared to welcome the survey, responding positively to the need for research on this topic and sometimes even sharing that they knew of cases, before being asked. This might suggest that respondents were less worried about disclosure, bearing in mind our question was phrased as “you or someone in your household”, than we had predicted. This is possibly due to the length of time since the war but equally may be due to the efforts of civil society organisations to challenge the stigma survivors of WSV face and campaigning for justice.⁶ Enumerators suggested that these respondents would have been happier to not self-complete as a result. In contrast, in areas with low documented experiences of WSV, enumerators suggested that respondents were happy to not self-complete possibly because they did not have any WSV experience. Hence there could be plausible reasons why respondents in both high and low prevalence regions might prefer the interviewer to complete the survey rather than opting for the privacy offered by self-completion. We certainly do observe a strong regional dimension in interview mode which could support this insight: we observe some municipalities with very high interviewer administered rates such as Pristina, Mitrovica, Pudujevo, Drenas and Vitina, and all of these have clusters of documented WSV.⁷ It is however difficult to think about a credible way to test the regional characteristics that are associated with lower self-completion, in particular given that we do not have information on prevalence of WSV that is granular enough to conduct such tests.

Interestingly, interviewers did not report issues with the format or formulation of the question. Arguably, item-list

⁶ There are numerous organisations campaigning for justice for survivors of WSV and their families. Of particular note are four organisations which together form the *Be My Voice* platform, supporting survivors of WSV through the complex process of applying for recognition and compensation: Kosovan Rehabilitation Centre for Torture Victims, Center for the Promotion Women Rights, Medica Kosova and Medica Gjakova, which provide psychological counselling, research and documentation of cases, and campaign for justice nationally and internationally.

⁷ The survey was piloted on a small sample in Prishtina but the issue of interview response mode did not arise then.

questions can be difficult to answer, and we may have added to the cognitive demand by focusing on experience beyond the respondent's own. As explained in our design section, asking about the household experience is also helpful in encouraging respondents to disclose a stigmatised experience without revealing it as being a personal experience and is thus frequently used in direct questions and indirect questions about sexual violence (e.g. Koos and Lindsey, 2022). While our interviewers' training session focused extensively on the handling of this item-count question, and on how to clarify the items and support respondents without interfering, the formulation of the question as including experience in the household did not generate questions during training or piloting. Ex-post, interviewers did not express concerns over cognitive load or report difficulties on answering this question. Our discussions with the enumerators also suggest that the decision not to pass on the tablet to the interviewee, or for the interviewee not to accept it, was made before the question was read rather than after. The form of the question itself was thus not a primary issue in that decision.

In summary, these focus group discussions revealed on the one hand that there should be individual characteristics of respondents that are relevant to the choice of interview mode, but also that more complex dynamics around the interaction between interviewer and respondents could have been at play. We conducted a focus group discussion but did not field a comprehensive enumerators' survey to assess their sense of what happened in the area they worked on as we felt it could be biased if done ex-post, and after we reported concerns over the extensive use of interviewers' help. We can note that interviewers were working in pairs in small locations (PSU) within their own municipalities, but away from their own settlements. This means that their ways of speaking were familiar to respondents (e.g. language, dialects, accents) and they were attuned to local sensitivity but also unknown to the people they interviewed. While the enumerators who took part in the focus group felt that interviewer completion was a natural and neutral choice given their local expertise and experience, it may not always have been the case.

3.5 Selection into interview mode and item count

Finally, we explore the drivers of interview-mode selection and the degree to which responses to our indirect item-count question may have been impacted by interview mode.

First, we explore whether there were characteristics of respondents systematically associated with interviewer-assisted completion. We estimate a linear probability model of the correlates of interview mode, with the dependent variable equal to 1 if the interview was interviewer-assisted and

0 if self-completed or family-assisted. We include a range of respondent characteristics such as age, gender, education, employment status, ethnicity and fixed effects for interviewer and municipality. Table A1 in the Supplementary Information shows the results. We find that male respondents, older respondents, those with no formal education, retired people and stay at home workers were all more likely to have received the assistance of the interviewer, while women, younger respondents, unemployed and ethnic Serbs were more likely to complete the interview themselves. This can be interpreted as broadly consistent with a position that some people will be less comfortable with the technology and others more worried about their verbal response being overheard and thus confirm some of the reasons given by enumerators to explain why some individuals preferred not to take the tablet.

However, the effects of work status or ethnicity mostly disappear when we add in interviewer fixed effects and when we control for the sample type. The interviewer fixed effects are also strongly predictive of the interview mode: they are jointly highly significant and strongly increase the model's R-square. In our view, this demonstrates that interviewer's choices shaped the interview mode above and beyond what we might expect from respondent characteristics and preferences alone. In a final specification, we also add an indicator of wartime experience. This variable takes value 1 if the respondents declared having experienced death or injury in the family during the conflict or displacement and thus does capture forms of victimisation that are typically captured in direct questions in post-conflict settings (Brück et al., 2016) and that are distinct from those listed in our count-item experiment. This dummy is insignificant, illustrating that interview mode selection does not seem to be driven by general victimisation. We cannot rule out however, that selection into interviewer-led responses was driven by other aspects of victimisation, trauma or grievances that we did not measure in our survey.

To explore further the association between interview mode and responses, we estimated an OLS model with the item count as the dependent variable, using the same set of respondent characteristics and adding in treatment, interview mode and an interaction term. Table 1 shows that where the interview mode is interviewer assisted, item counts are higher, and statistically significantly so, than when self-completed. This is true even when we include controls for the characteristics of the participants and fixed effects for municipalities. This effect is thus not explained away by observables. Additionally, when we add an interaction effect between interview mode and treatment, we find the coefficient for the interview mode to be mostly unchanged, while the interaction term is insignificant. This suggests that the interview mode effect is mostly carried through the non-treated respondents. Thus overall, when-

Table 1*Drivers of item count: treatment and interviewer mode*

	Model 1 OLS		Model 2 OLS		Model 3 OLS		Model 4 Poisson	
	Estimate	Standard error	Estimate	Standard error	Estimate	Standard error	Estimate	Standard error
Treatment (=1 if long list, 0 otherwise)	-0.055	0.059	-0.002	0.077	-0.005	0.06	-0.001	0.027
Interview mode (=1 if interviewer assisted, 0 otherwise)	0.180**	0.059	0.238**	0.081	0.175*	0.082	0.103**	0.04
Treatment x Interviewer-mode interaction			-0.12	0.119	0.008	0.086	0.001	0.036
Municipality FE					Yes***		Yes***	
Interviewer FE					Yes***		Yes***	
Individual controls (as in model 3, Supplementary Information)					Yes***		Yes***	
Constant	2.233***	0.048	2.208***	0.053	2.890***	0.374	1.071***	0.155
Observations	1573		1573		1573		1573	
R ² (Pseudo-R ² col 4)	0.006		0.007		0.503		0.097	

Cells display regression coefficients and robust standard errors. Columns 1–3 are OLS estimates, Column 4 are results from a Poisson model

* $p < .10$, ** $p < .05$, *** $p < .01$

ever the interviewers filled the indirect item count response themselves, the reported count was higher, but especially so for the items presented in our short list. The size of the coefficient also shows that the interviewer's impact on the reported count was much larger than known incidence rate of WSV, making it nontrivial in our context. As a robustness check, we also ran a Poisson regression model with the same specification with qualitatively similar results, shown in column 4.⁸

The literature has documented cases where respondents appeared to “flee” from reporting stigmatised experiences by biasing their answers to item count question downward when a sensitive item was included (Chuang et al., 2021, Kotsadam and Løvgren, 2025). While this might be at play here, the over-riding effect detected is one by which respondents report systematically more varied aspects of victimisation when responding to our item count question openly (via the interviewer) rather than more privately. Importantly, this is the case even though other war time experiences reported through direct questions appear irrelevant to interview-mode selection.

The mode of interview is thus relevant to the responses obtained. This could translate a non-random selection of respondents into the interview mode beyond what we are able to test based on the information we collected and/or

a situation whereby the interviewers have inadvertently impacted on the responses received. Our sample was selected to be sufficient on the basis of (i) a possibly large under-reporting, or a large effect of WSV on our secondary outcome and (ii) a simple treatment. It is however too small to carry out a detailed and meaningful analysis of the sub-samples created by the break-down of the interview protocol. We do not have enough information and data to identify clearly why respondents have selected (or been selected) into a specific interview mode and how it relates to their item count response.

4 Conclusions

This paper reports on a list experiment designed to measure the prevalence of war time sexual violence in Kosovo. The aim of the experiment was to investigate and quantify a potential bias in WSV reporting in Kosovo, as in the presence of strong social stigma, not all victims/survivors would be expected to report their experience in a direct question. While in our application, the use of an indirect question does not reveal that answers to a direct question reflect an under-reporting of WSV, a more detailed exploration of the data suggests other possible biases.

Our experiment design is consistent with examples from the existing literature using list experiments to measure WSV and our presentation of it in this paper should provide some insights for those seeking to design and implement similar item-count questions regarding key issues to consider. A key assumption for the validity of the method is

⁸ Note we also estimated models 3 and 4 for the main and supplementary samples separately. Treatment remains statistically insignificant throughout, as does the interaction term. Interview mode remains statistically significant at the 10% level but only for the Poisson estimates for the main model.

that respondents are truthful in their response—something that cannot be tested formally. Here the patterns of response from interviewees appear to differ depending on the interview mode: whether the interviewer stepped in to help the respondents fill their answers or not. Therefore, instead of having a well-defined and exogenous treatment (being exposed to a long or short list of items), we instead have several variations reflecting whether responses to our sensitive question were reported privately. And these variations were not randomly allocated.

We observe that there is a different pattern in responses between interviews with the indirect question self-completed or with family assistance compared to those completed by the interviewer. Specifically, even after controlling for personal characteristics, those who were helped by the interviewers to enter their responses consistently report to have suffered from a wider range of war experiences (i.e. a higher item count), except when one of these experiences is WSV. This implies that the mode of completion interfered with our design and suggest complex social and psychological processes at play. On the one hand, the consistent over-reporting we identify could suggest that respondents with stronger experience of the war want this experience recognised and are more likely to self-select into being helped, or that being helped itself triggers a greater propensity to report war experiences. On the other hand, a relatively lower reporting when WSV is mentioned could suggest on the contrary a possible fleeing behaviour. In practice, we can assign some of the interview mode variation to characteristics of individuals which might make them less confident in using technology, such as being older, with less formal education, retired or stay at home carers, but much of the variation appears to be captured by interviewer fixed effects.

Interviewers suggested two competing reasons for why some respondents chose not to self-complete the survey: that some respondents welcomed the survey, particularly those in areas believed to have been the sites of WSV and even expressed a willingness to share more about their experiences, while in other areas with lower documented cases of WSV, respondents seemed to dismiss the need to self-complete, apparently not having anything they feared disclosing openly. However, this reluctance to self-complete may itself be a protective behaviour. The stigma associated with WSV in Kosovo is believed to be very strong. Only three survivors⁹ have made public statements about what happened to them, attracting both respect and derision from different parts of society, and many survivors prefer anonymity.

We believe that, given the relative paucity of similar studies, our research provides useful lessons for other researchers seeking to study wartime sexual violence or to use item-count questions on other sensitive experiences. Based on this experience, we would first consider several changes in our original approach to preparing for the survey. To start with, we would address potential failings in our training of enumerators. As the list experiment was a new method for the team, we perhaps spent a disproportionate amount of time on the questions themselves rather than the mechanisms to be used for self-completion. Dedicating time to explaining the importance of handling the question consistently over the entire sample of respondents might help offset some of the behaviours we observed among the interviewers. Our interviewers felt that helping respondents to fill the sensitive questions on the tablet was a neutral decision with no impact on responses, but the data shows otherwise. Using a much larger sample could also have helped mitigating the effect of this unplanned variation in implementation. Next, we would also conduct a larger and more exhaustive piloting of the survey instrument, experimenting with different methods to address the ethical issues, including offering an alternative way to complete the sensitive questions, such as filling in a form and placing the response in an envelope, or handing respondents a card to read. Lastly, we would consider an incentive structure for enumerators which rewards self-completion of the sensitive questions, and to respondents to thank them for their time and the use of a larger sample size.

However, in our view, the primary lesson here is that aside from the design issues in any survey experiment, not least an item count experiment, a significant challenge arises from the social context in which the study takes place. In Kosovo, over 20 years after the conflict, it appears that reporting WSV is still sensitive enough that respondents may actively bias their responses to a non-intrusive question to make sure they are not believed to have suffered from it, a behaviour that has been described as “fleeing” in earlier work (Chuang et al., 2021). At the same time, there may still be a strong need for a population that suffered greatly during the conflict to get their experience recognised, since we identified an unexpected higher reporting of other experiences of war victimisation, when these experiences are recorded by an interviewer, rather than more privately. Our work suggests complex social desirability biases in the report of varied aspects of war victimisation overall in Kosovo and illustrates the high sensitivity of the item-count method to small changes in implementation.

Acknowledgements The data used in this study was collected thanks to a British Academy Small Research Grant and additional funding from the Department of Economics at University of Sussex. Data collection was conducted by UBO Consulting and our particular thanks go to Hana Bacaj and Rron Lecaj for leading this. The authors also ac-

⁹ Vassije Krasniqi Goodman, Shyhrete Tahiri Sylejmani and Ramadan Nishori are all working to raise awareness, reduce stigma and support survivors.

knowledge financial support from CNET (SSEES, UCL) at the development stage of this project. Helpful feedback on the research design was received from the participants of a workshop funded by CNET in July 2023. We are also grateful to colleagues who provided feedback on the survey design, including Alberto Prati and Vjollca Krasniqi, and to two anonymous reviewers and the guest editors for suggestions on earlier versions of this paper. We thank Fathir Suhaimi for research assistance during the summer of 2024. Ethical approval was secured from the University of Sussex in September 2023.

References

- Agüero, J., & Frisancho, V. (2022). Measuring Violence Against Women with Experimental methods. *Economic Development and Cultural Change*, 70(4), 1565–1590.
- Ahlquist, J. S. (2018). ‘List Experiment Design, Non-Strategic Respondent Error, and Item Count Technique Estimators’. *Political Analysis*, 26(1), 34–53.
- Amini, C., & Douarin, E. (2020). Corruption and Life Satisfaction in Transition: Is Corruption a Social Norm in Eastern Europe? *Social Indicators Research* 151, 723–766. <https://doi.org/10.1007/s11205-020-02389-6>.
- Aronow, P. M., Coppock, A., Crawford, F. W., & Green, D. P. (2015). “Combining List Experiment and Direct Question Estimates of Sensitive Behavior Prevalence”. *Journal of Survey Statistics and Methodology*, 3(1), 43–66.
- Asadullah, M. N., De Cao, E., Khatoon, F. Z., & Siddique, Z. (2021). Measuring Gender Attitudes using List Experiments. *Journal of Population Economics*, 34, 367–400.
- Bauer, M., Blattman, C., Chytilová, J., Henrich, J., Miguel, E., & Mitts, T. (2016). “Can War Foster Cooperation?”. *Journal of Economic Perspectives*, 30(3), 249–274.
- Blair, G., & Imai, K. (2012). Statistical Analysis of List Experiments. *Political Analysis*, 20(1), 47–77.
- Blair, G., Coppock, A., & Moor, M. (2020). When to Worry about Sensitivity Bias: A Social Reference Theory and Evidence from 30 Years of List Experiments. *American Political Science Review*, 114(4), 1297–1315.
- Blattman, C. (2009). From Violence to Voting: War and Political Participation in Uganda. *American Political Science Review*, 103, 231–247.
- Brück, T., Justino, P., Verwimp, P., Avdeenko, A., & Tedesco, A. (2016). Measuring Violent Conflict in Micro-Level Surveys: Current Practices and Methodological Challenges. *The World Bank Research Observer*, 31(1), 29–58.
- Chuang, E., Dupas, P., Huillery, E., & Seban, J. (2021). Sex, Lies, and Measurement: Consistency Tests for Indirect Response Survey Methods. *Journal of Development Economics* 108: 102582. <https://doi.org/10.1016/j.jdeveco.2020.102582>.
- Dautry, M. (2019). “The Endless Struggle of Wartime Rape Victims in Kosovo” Equal Times.
- Detkova, P., Tkachenko, A., & Yakovlev, A. (2021). Gender Heterogeneity of Bureaucrats in Attitude to Corruption: Evidence from List Experiments. *Journal of Economic Behaviour & Organization*, 189, 217–233.
- Douarin, E., Litchfield, J., & Sabates-Wheeler, R. (2012). War Legacies, Welfare and Livelihoods in Rural Areas: The Case of Kosovo. *Journal of Agricultural Economics*, 63, 201–222.
- Euronews (2022). “Why are Victims of Sexual Violence in Kosovo being rejected by the Commission?” Euronews Albania.
- European Commission (EC) (1999). Kosovo – Emergency Assessment of Damaged Housing and Local/Village Infrastructure. Available from: Final report4pdf.PDF (reliefweb.int) [last accessed 14 May 2021].
- Freedman, S. (2024). Preservation of Cultural Sites and Artifacts During Wartimes. *The Serials Librarian*, 85(1–4), 32–39. <https://doi.org/10.1080/0361526X.2024.2318219>.
- Glynn, A. (2013). What Can We Learn with Statistical Truth Serum? Design and Analysis of the List Experiment. *Public Opinion Quarterly*, 77(S1), 159–172. <https://doi.org/10.1093/poq/nfs070>.
- González, B., & Traunmüller, R. (2023). The Political Consequences of Wartime Sexual Violence: Evidence from a List Experiment. *Journal of Peace Research*, 8, 1–16.
- Haxhijaj, S. (2021). *The Enduring Agony of Wartime Rape in Kosovo* Votra Magazine
- Human Rights Watch (HRW) (2000). “Kosovo: Rape as a Weapon of ‘Ethnic Cleansing’”. <https://www.hrw.org/report/2000/03/01/kosovo-rape-weapon-ethnic-cleansing>. Accessed 29 July 2026.
- Igarashi, A., & Nagayoshi, K. (2022) Norms to be Prejudiced: List Experiments on Attitudes Towards Immigrants in Japan. *Social Science Research*, 102, 102647. <https://doi.org/10.1016/j.ssresearch.2021.102647>.
- Imai, K. (2011). Multivariate Regression Analysis for the Item Count Technique. *Journal of the American Statistical Association*, 106(494), 407–416.
- Iraman, E., Ono, Y., & Kakinaka, M. (2022). Tax Compliance and Social Desirability Bias of Taxpayers: Experimental Evidence from Indonesia. *Journal of Public Policy*, 42(1), 92–109.

- Justino, P., Brück, T., & Verwimp, P. (2013). *A Micro-Level Perspective on the Dynamics of Conflict, Violence and Development*. Oxford University Press.
- Koos, C. (2018). Decay or Resilience?: The Long-Term Social Consequences of Conflict-Related Sexual Violence in Sierra Leone. *World Politics*, 70(2), 194–238.
- Koos, C., & Lindsey, S. (2022). Wartime Sexual Violence, Social Stigmatization and Humanitarian Aid: Survey Evidence from Eastern Democratic Republic of Congo. *Journal of Conflict Resolution*, 66(6), 1037–1065.
- Koos, C., & Traunmüller, R. (2022). *The Social and Political Consequences of Wartime Sexual Violence: New Evidence from List Experiments in Three Conflict-affected Populations*. WIDER Working Paper 2022/11. Helsinki: UNU-WIDER
- Kosovo Centre for Rehabilitation of Trauma Victims (KRCT) (2021) Beyond Pain Towards Courage: Stories about the Trauma of Wartime Sexual Violence, Pristina, Kosovo.
- Kosovo Centre for Rehabilitation of Trauma Victims (KRCT) (2023). “Documentation of Sexual Violence Crimes during the War in Kosovo”. <https://krct.org/dokumentimi/en/documentation/>. Accessed 29 July 2026.
- Kotsadam, A., & Løvgren, M. (2025). Is it Time to Put a Moratorium on List Experiments for Domestic Violence Elicitation?. *Sociological Methods & Research*, 00491241251372542.
- di Lellio, A., Sada, M., & Kraja, G. (2018). “Wartime Sexual Violence: New Evidence as Hundreds of Survivors come Forward”. Accessed 20 Jan 2024. Prishtina Insights.
- Li, J., & Van den Noortgate, W. (2022). A Meta-analysis of the Relative Effectiveness of the Item Count Technique Compared to Direct Questioning. *Sociological Methods & Research*, 51(2), 760–799.
- Litchfield, J., Douarin, E., & Gashi, F. (2024). Angry Men and Civic Women? Gendered Effects of Conflict on Political Participation in Kosovo. *Feminist Economics*, 30(2), 257–296.
- Organisation for Security and Co-operation in Europe , OSCE (1999). “*Human Rights in Kosovo: As Seen, As Told*”. Volume I, October 1998 - June 1999. Organization for Security and Co-operation in Europe.
- Stancheva, S. (2023). “How to Run Surveys: A Guide to Creating your Identifying Variation and Revealing the Invisible”. *Annual Review of Economics*, 15(1), 205–234.
- Trako. I. (2018) “Returning Home After Conflict Displacement: Labor Supply and Schooling Outcomes Among Kosovar Households.” *HAL-SHS Sciences de l’Homme et de la Societe* halshs-01828477. <https://url.uk.m.mimecastprotect.com/s/uoW2CzXxRcMDmgqjt4f7js9rTsv?domain=shs.hal.science/>.
- Traunmüller, R., Kijewski, S., & Freitag, M. (2019). The Silent Victims of Sexual Violence during War: Evidence from a List Experiment in Sri Lanka. *Journal of Conflict Resolution*, 63(9), 2015–2042.
- Tsai, C. (2019). Statistical Analysis of the Item-count Technique using STATA. *The Stata Journal*, 19(2), 390–434.
- Tsuchiya, T., Hirai, Y., & Ono, S. (2007). A Study of the Properties of the Item Count Technique. *Public Opinion Quarterly*, 71(2), 253–272. <https://doi.org/10.1093/poq/nfm012>.
- United Nations Secretary General (2019). “*Conflict-Related Sexual Violence Report*” S/2019/280.
- Zaal, E., Ongena, Y., van der Velden, N., Loughnan, D., & Hoeks, J. (2026). Unraveling Honest Responding: A Systematic Review on the Effectiveness of Social Desirability Bias Reduction Methods in Survey Research. *Quality & Quantity*, 60(3), 10359–10391.