IMPACT Minimum Standards Checklist for Semi-Structured (Qualitative) Data Processing and Analysis

This checklist aims at summarising all the necessary documents and steps needed to achieve a fast and thorough validation of the processing and analysis of semi-structured, qualitative data.

1. Key Documentation Needed for Analysis Validation by HQ

The following list contains the documentation needed by the HQ Research Design and Data Unit (RDDU) to validate data processing and analyses performed on semi-structured, qualitative data.

➢ **<u>Always shared for HQ review</u>:**

1. **Enumerator debriefs** – After every data collection activity (interview/discussion), whether qualitative or quantitative, teams must debrief with enumerators. In these debriefs, enumerators are asked to share their thoughts and experience on the data collection activity, as neutrally as possible. Enumerator input should complement results. In cases where an enumerator observed something that contradicts a response, this should be allowed to be acknowledged in the analysis, if even just to flag a risk of respondent bias. Summaries of these debriefs must be recorded and shared with HQ. Enumerator debriefing does not require a specific form, and so summaries can be shared in any format for HQ review. However, some examples of existing enumerator debrief forms can be found here: Word format example and Excel format example. These debriefs must be anonymised and shared with the RDDU for verification and saving to the server.

2. **Finalised transcripts** – Once enumerator debrief summaries, handwritten notes and interview/discussion recordings (if available) collected during the data collection activity have been gathered by the team, these must be used to develop anonymised and translated transcript, in a script-like, verbatim manner must be developed. This transcript should be in either English, French or Spanish (depending on the context of data collection) and should contain as much content as possible, including non-verbal responses[1] that have been recorded. Ultimately these files should be used for data processing and analysis, and should be shared with the RDDU to assist their review and validation of the semi-structured, qualitative data processing and analysis.

3. **Complete Data Saturation and Analysis Grid** – During semi-structured, qualitative data collection, teams must use the Data Saturation and Analysis Grid template available here on the Repository. After each interview/discussion is completed, a new column must be added to the saturation grid, recording all discussion topics and points. This means that the Grid must be developed on an ongoing and continuous basis throughout qualitative data collection. When a new point is raised in the latest transcript, a new row must be added. Ultimately, data saturation has been achieved when no new rows are added for a particular stratum, at which point data collection can stop for said stratum.[2] Once all strata have been completed, the total number of references for each discussion point can then be tallied and a summary

[1] Non-verbal responses are any actions or inputs presented by respondents that cannot be captured by the words they shared. These could include but are not limited to general behaviour, gestures, tone, postures, pauses or emotions.
[2] Achieving data saturation is the ideal means by which semi-structured, qualitative data collection can stop. However, it should be reiterated that in many cases, the number of interviews conducted is based on capacity and resources.

of key findings noted in the file, following the template. This must be shared with the RDDU for data processing and analysis validation.

4. **Compete Method Report** (integrated in the Data Saturation and Analysis Grid template) – The Method Report template can be found in the Data Saturation and Analysis Grid template here. This consists of a brief recap of the data collection and analysis methods used, the choices that were made throughout the analysis process, including their justifications, strengths and limitations, as well as a publication plan for the semi-structured, qualitative analysis. The justification for including a Method Report builds upon ideas identified during literature reviews, where the importance of having clear and transparent qualitative analysis is made essential.[3] Accordingly, anyone should be able to view the Method Report, take the input transcripts, analyse by following the same process and ultimately reach the same or similar conclusions, improving analytical transparency. This must be shared with the RDDU at the same time as the Data Saturation and Analysis Grid (using the available template) so that they can verify the analytical process followed.

➢ **Shared for HQ review if applicable:**

1. **Extended Data Analysis Plan (DAP)** – Semi-structured, qualitative data processing and analysis must always be conducted in line with the original DAP, validated at the Research Design stage. If changes have occurred during the implementation of data collection[4] and/or if the analysis has been extended in comparison to the original DAP, an extended DAP must be prepared, where the team must build upon the original DAP, highlighting changes and tracking additions to the analysis plan. The DAP template can be found in the Data Analysis Plan (With Example) template in the online Repository toolkit.

2. **Further Analysis Files** – After the Data Saturation and Analysis Grid has been developed, if the assessment requires further analysis and team capacity allows, further analysis files can be produced. This can consist of analysis using analytical software such as NVivo or Atlas. These can be produced to complement the Data Saturation and Analysis Grid, but it cannot replace the grid. The Data Saturation and Analysis Grid and further analysis files must speak to each other, with the further analysis codebook being produced to correspond with the discussion topics and points identified in the Grid. These further analysis files can be shared in their original format (or exported format, if available), with the input files and subsequent coding included for RDDU review and validation.

2. Preferred Tools

Available Resources

A range of resources and templates are already available on the Repository to support planning and preparations of qualitative data processing and analysis:

- Data Analysis Plan template – This template must be used when preparing the Data Analysis Plan during the research design phase of all projects, regardless of whether they draw on quantitative or qualitative methodologies.

[3] Morvcsik, A. *Transparency: The Revolution in Qualitative Research*, American Political Science Association, 2014 and Aguinis, H & Solarino, A. M. *Transparency and Replicability in Qualitative Research: The Case of Interviews with Elite Informants*, Strategic Management Journal, 2019.
[4] Note here that changes can to the analysis plan can only be made after the data collection and analysis process is underway if they remain in line with the pre-defined research objectives and research questions outlined in the TOR. Changes beyond the scope of the assessment are not permitted.

- Enumerator Debrief Form examples – While it is required that teams conduct enumerator debriefs and that these debriefs are noted in some format, it is not an obligation to use a format or template. To give some examples of enumerator debrief templates used in some countries, see the following:
  - Word format example
  - Excel format example
- Data Saturation and Analysis grid template (Updated template published October 2020) – This template must be used to process qualitative data, i.e. monitor data saturation, throughout data collection, and to conduct analysis by summarizing key findings. In addition, the Method Report sheet must be filled in before submitting the file for HQ review and validation, to provide transparency on the qualitative analysis process.
- READ_ME Sheet guidelines – This can be used to develop a detailed READ_ME sheet to accompany the Data Saturation and Analysis Grid. A READ_ME can be included if the analysis is complicated or extensive, to help the RDDU navigate the document during the review process, or must be included if the grid is to be published, to help wider audiences understand the approach.

Analysis Software

When completing the Data Saturation and Analysis Grid, the Excel template must be used. However, for additional analysis, different softwares can be used and as with any IMPACT research and analysis, there is flexibility here. In all cases, teams are welcome to use Excel, Nvivo or Atlas for further analysis.

If teams would like to use any other software for analysis, then they should reach out to the RDDU to discuss, before beginning analysis to ensure it will also be possible on HQ's side to review the analysis in this format. Regardless of the software used for further analysis, codebooks/coding structures should directly correspond to the saturation grid's discussion points, to ensure analysis is consistent.

3. Summary of Minimum Standards

The table below summarises minimum standards that IMPACT aims to fulfil for the processing and analysis of qualitative data. The table is to be read as follows:

- The Standard is an analysis principle that is to be followed in qualitative analysis approaches within IMPACT, based on principles commonly followed in qualitative research.
- The Requirement column defines those that should be considered minimum standards. All the Required standards need to be accomplished by all qualitative data processing and analysis across the five mentioned assessment phases, while in some cases standards are flagged to be completed where the team has capacity or where the requirement is relevant for the assessment.

Minimum standard	**Requirement** (September 2020)
Analysis Plan / Research Design	
Qualitative data processing and analysis planning must be clearly detailed in the Terms of Reference (TOR) and Data Analysis Plan (DAP) developed during the Research Design phase This can be achieved simply by following the TOR and DAP templates found in the Repository toolbox. Note that the DAP should include clear research questions in line with assessment objectives and include intentions to aggregate findings and identify differences between respondent groups. If any changes are made to the DAP during the analysis process, an Extended DAP must be	Required

developed and shared with HQ, with changes and additions to the original DAP clearly highlighted.[5]	
The Data Saturation and Analysis Grid template should be reviewed and understood by the team in advance of data collection The team must download the template, which contains both the Method Report and the Data Saturation and Analysis Grid, from the Repository. The team must ensure they are clear on which sections must be filled out during and after data collection. Note that discussion points recorded in the Grid must be specific and must represent the content of the discussion point; it cannot be vague or overarching in content.[6] If there are any questions, the team must reach out to the RDD Unit before beginning data collection.	Required
Data Processing, Checking, and Cleaning	
Audio recordings of interviews should be used where feasible and appropriate, and when consent to record has been provided, to help ensure accurate transcription If audio recordings have been collected, they must be stored offline in a password protected file, and must be destroyed once full transcripts have been developed	Done where capacity
Enumerator debriefs must be conducted with the enumerator as soon as possible after the interview/discussion has been completed Enumerator debriefs should complement any handwritten notes recorded throughout the interview to generate final translated transcripts. Summaries of enumerator debriefs must be anonymised before storing. These debriefs can be summarised in any format the team chooses, although these summaries must be shared with HQ for verification. Some examples of enumerator debriefs can be found here in Word format or in Excel format.	Required
Comprehensive transcript packages for each and every data collection exercise (every interview/discussion) must be produced, developed using all available resources (audio recording, field notes, enumerator debriefs and any other available resources) Transcripts must: - Follow a verbatim, script-like structure - Capture non-verbal responses - Be translated to the most accessible language for the mission (e.g. English, French or Spanish) - Be anonymised and have any sensitive information removed before sharing and storing.	Required
Data collection and data saturation must be tracked daily throughout the data collection phase This includes monitoring population group and geographical coverage to ensure either data saturation has been achieved[7] and/or the original sampling strategy has been fulfilled. Ideally qualitative data collection would be guided by data saturation. However, if resources do not allow for further interviews in order to meet data saturation, then data collection must stop regardless. If the minimum number of interviews outlined in the Research Design phase has not been achieved, the reasons and consequences of this must be specified in the Method Report (found in the Data Saturation and Analysis Grid template), and the team must reach out to the RDDU to discuss options as there is the possibility that there is insufficient data from which to draw findings.	Required
Before finalising files and sharing with the HQ RDDU, all files (including final transcripts and enumerator debrief summaries) must be checked and validated in-Country These checks need to ensure all personal data has been anonymised, sensitive information has been removed and all necessary metadata (number of participants, gender of participants, date and location of data collection etc) has been included. In addition, review team members in-country (such as Research Managers or Country Coordinators) ensure the preparation of all files are clear	Required

[5] Note here that changes to the analysis plan can only be made after the data collection and analysis process is underway if they remain in line with the pre-defined research objectives and research questions in the TOR. Changes beyond the scope of the assessment are not permitted.
[6] Specifically, this means that discussion points should not broadly refer to the theme of discussion (e.g. Food Security) but should rather reflect the detailed content of the point raised (e.g. Food insecurity- Not enough food; Food insecurity- Poor quality food; etc).
[7] Data saturation is met when, for a particular stratum, no new discussion topics or points are raised in the latest interview or discussion. In practice, provided the Data Saturation and Analysis Grid is updated throughout data collection, this means that for a particular stratum, no new rows were added, because no new discussion points were raised in the newest interview added to the grid. In this case,

and coherent, ready for review.	
Implementation of Analysis	
Analysis must be guided by the original TOR and DAP, or by the Extended DAP if needed The TOR/DAP should be followed as a minimum. However, additional analysis is permitted provided the team explain clearly in the Extended DAP why this additional analysis is important and how it continues to fulfil the research questions	Required
Data processing and analysis must be conducted using the Data Saturation and Analysis Grid template first, followed by further analysis if relevant The data processing and analysis steps taken must be transparent and clearly documented in the Method Report in the Data Saturation and Analysis Grid template. All stages of any further analysis conducted must also be clearly outlined to ensure transparency.	Required
The analytical process must be as neutral and unbiased as possible It can be tempting to make a lot of snap decisions while processing and analysing qualitative data on what is and is not relevant. However, this tends to bias the results, as this leads to key findings being omitted based on the analyst's own judgement. Effort should be made to remain neutral and analyse comprehensively, using the pre-outlined research questions from the TOR to guide this process.	Required
The analytical process must be consistent, meaning teams should not change the analysis approach halfway through Provided the data processing and analysis plan outlined in the TOR and (Extended) DAP are followed, this should not happen. In rare and specific cases there can be a specific need to change the approach. If this is the case, teams must reach out, discuss and agree on this alteration with the RDDU in advance of making any change.	Required
Analysis should be developed by drawing predominantly on verbal responses but should also factor in non-verbal responses. Where possible, enumerators should be trained to include non-verbal responses (such as sighing, tone, body language, pauses and emotion etc) into their field notes. These inputs from enumerators must never negate or impact our analysis of verbal responses, as these inputs are subjective, but they can be used to build a bigger picture and strengthen how verbal responses are interpreted.	Done where capacity[8]
Before finalising files and sharing with the HQ RDDU, all files (including the Data Saturation and Analysis Grid) must be checked and validated in-Country These checks need to ensure all personal data has been anonymised, sensitive information has been removed and all necessary metadata (number of participants, gender of participants, date and location of data collection etc) has been included. In addition, review team members in-country (such as Research Managers or Country Coordinators) ensure the approach to and content of the analysis is in line with the overall purpose of the research.	Required
File Sharing for HQ Validation	
See Section 1 above	Required
Qualitative Data and Analysis Publication	
It is strongly encouraged that teams upload the Data Saturation and Analysis Grid to the Repository (especially from REACH research cycles), for publication This step vastly improves the transparency of qualitative analysis and subsequent outputs that IMPACT is producing. However, it is understood that in some cases this might not be possible due to the nature of the project (e.g. for Third Party Monitoring) or because data will be too sensitive to allow this. In this case, the Data Saturation and Analysis Grid should be archived on the Repository, for internal reference. In order to publish the Data Saturation and Analysis Grid, the file must be tidy and must include a READ_ME sheet to explain how to navigate the document. Only validated files can be uploaded to the Repository or be shared externally.	Done where appropriate

[8] To clarify, it is a requirement that analysis uses verbal responses. Non-verbal responses are to be incorporated into analysis only where there is capacity.