

# Integrity Watch Afghanistan Research Methods Handbook

February 9 2021

David Carment, S. Yaqub Ibrahimi, Daphne Wang, Maseh Hadaf  
Ottawa, Canada

@NPSIA-CIFP

# Table of Contents

Table of Contents .....	1
Table of Tables .....	3
Table of Figures .....	4
Introduction .....	5
Part 1: Research and Evaluation Design .....	6
Research Design .....	6
1. Statement of the Research Topic and Objectives .....	6
2. Key Questions/Assumptions .....	6
3. Brief Explanation of Rationale and Significance .....	6
4. Methodology and Design .....	7
5. References .....	15
Needs Analysis .....	15
Evaluation Design .....	16
Pre-Assessment Analyses .....	17
Results Based Monitoring and Evaluation Design: Descriptive and Normative Questions ..	25
Impact Evaluation Questions .....	28
Addressing Impact-Related Questions Through Process-Based Approaches: .....	29
Participatory Approaches .....	31
Peace and Conflict Impact Assessment (PCIA) .....	32
Managing Evaluations .....	36
Ethical Considerations .....	37
Creating an Accountability Framework .....	37
Ethical Considerations for Research Design .....	38
Recognizing Research Corruption .....	40
Part 2: Methodologies for Research and Evaluation .....	41
Sampling .....	41
Surveys .....	47
Focus Groups .....	58
Part 3: Research and Evaluation Presentation .....	68
Policy Briefs and Policy Papers .....	68
Innovative Tools .....	85

References .....	89
Appendix A: Checklist for creating a good TOC .....	98
Appendix B: Optimizing quasi-experimental approaches: .....	99
Appendix C: Alternatives to establishing causation .....	103
Appendix D: Cost-effectiveness analysis .....	104
Appendix E: Focus Group Consent Form Example .....	105
Appendix F: Template for an Evaluation Term of Reference .....	107

# Table of Tables

Table 1: Variable-based research designs .....	10
Table 2: Tools for Bridging the Qualitative-Quantitative Divide .....	14
Table 3: Conflict Analysis Template Drawn from the Country Indicators for Foreign Policy (CIFP) Fragile States Methodology .....	18
Table 4: Conflict Analysis Scenarios template .....	20
Table 5: Profiling stakeholders .....	22
Table 6: Sufficient vs Necessary Conditions .....	30
Table 7: Stratified Sample of 30 MPs of Afghanistan's Nine Major Provinces .....	44
Table 8: Respondents' support for the post-2001 Afghanistan political system .....	56
Table 9: Canadians' responses towards federal elections.....	56
Table 10: Policy documents and their use cases .....	69
Table 11: Stakeholder checklist .....	80
Table 12: Strength of Causal Inference.....	103

# Table of Figures

Figure 1: Example of independent/dependent variables .....	7
Figure 2: Intervening Variables example .....	8
Figure 3: Theory of Change components.....	23
Figure 4: Theory of Change example .....	24
Figure 5: Interrelated Levels of conflict analysis .....	33
Figure 6: Selection Criteria Equation .....	45
Figure 7: Nominal question examples .....	52
Figure 8: Ordinal question examples .....	53
Figure 9: Interval question examples.....	53
Figure 10: Bivariate analysis of survey data .....	57
Figure 11: Descriptive code .....	62
Figure 12: In Vivo codes .....	62
Figure 13: Eclectic codes.....	63
Figure 14: From Codes to Category .....	64
Figure 15: From codes to results/theory .....	65
Figure 16: Table of contents in the Microsoft Word ribbon.....	73
Figure 17: Example of a box.....	79
Figure 18: Difference in Difference assessments .....	100
Figure 19: Non-Equivalent Comparison.....	102

# Introduction

This handbook provides instructions on how to conduct policy-oriented research as it pertains to the work of Integrity Watch Afghanistan (IWA). The handbook is designed as a practical reference for researchers in the field, covering the *how*, *what*, and *when* of policy-oriented research. It is neither comprehensive in scope nor does it dwell on the theories underpinning its recommendations. References are included for detailed reading.

The credibility of IWA's policy recommendations rests on sound research design and implementation. The handbook is structured accordingly and will guide the reader through three steps:

- 1) Research, needs analysis, and evaluation design;
- 2) Implementation of research and evaluation methodologies; and
- 3) Research and evaluation presentation.

Part 1 explains three different designs: research needs analysis, and evaluation design. The research design section discusses how to establish a research purpose and research objectives using IWA's format for a research proposal. These objectives are then tied to a suitable conceptual framework, and these concepts are then operationalized into observable or measurable variables. Here there is a discussion on internal validity, variable-based research designs, and their use cases. The needs analysis section includes its purpose and basic steps. Finally, the evaluation design provides pre-assessment analyses, questions, and approaches for answering result-based as well as impact evaluations. A section for peace and conflict impact analysis is included as a special kind of evaluation tool. What follows is an overview of general ethical considerations and research standards to be applied to the study.

Part 2 focuses on the implementation of research and evaluation methodologies. The focus is on three data collection methods: sampling, conducting surveys, and focus groups. Practical considerations including data collection and managing evaluations are also discussed.

Part 3, outlines research presentation concepts with a focus on policy-oriented report writing. This section identifies different types of policy briefs with their use-cases, a section-by-section guide on how to write a policy paper, guidelines on citations and referencing using the Chicago Manual of Style 17<sup>th</sup> edition Notes-Bibliography (NB) system, and innovative tools to support the research and report-writing process.

Beyond an initial reading of part 1 on research design, this handbook is meant to be a reference and does not need to be read in chronological order. Rather, it should be consulted as and when needed throughout the course of a research project.

# Part 1: Research and Evaluation Design

## Research Design

Research design is the basis for any research project. IWA recommends the following structure for a research proposal to outline your research design:

1. Statement of the Research Topic and Objectives
2. Key Questions/Assumptions
3. Brief Explanation of Rationale and Significance
4. Methodology and Design
5. References

The completion of this outline ensures that sound research principles are applied from the start and that the implementation and presentation of research rests on a rigorous foundation. Each section and sub-section is described below.

### 1. Statement of the Research Topic and Objectives

The first step is to establish a purpose and outline the objectives of the research to be undertaken. It is important to be clear and specific, rather than comprehensive. In one or two brief paragraphs, state the issue that you will focus on, followed by the explicit objectives of your proposed research.

### 2. Key Questions/Assumptions

What question(s), as related to your objectives, are you trying to answer? Briefly state: *In order to reach X objective and get answers for Y question, I plan to conduct research on Z topic.*

Aim to explore a subject in a way that might shed new light on issues that you are examining through your research—including a literature review, experts' lectures, field visits, and talking to people.

Refer to subsequent sections on research methodology on how to create research questions.

### 3. Brief Explanation of Rationale and Significance

What is the significance and contribution of your project to research to Afghanistan, a conflict-affected country, for policy makers and practitioners, and for human security and human development? In one paragraph, show how your research relates to and will advance policymaking or policy debates, if applicable. Keep in mind that your rationale must reflect the priorities of IWA and be timely, action-oriented, and considerate of relevant stakeholders.

## 4. Methodology and Design

### Establishing a conceptual framework

Begin by trying to identify a conceptual framework from past literature that can be applied to your research problem. You should make it clear how your research builds on past research (i.e. fills a gap, tests an explanation, answers a question raised by a past study) and how your problem statement relates to the proposed conceptual framework.(Miller and Salkind 2002) The conceptual framework will identify key concepts for your topic of interest and offer an explanation with predictive power to explain how those concepts interact. It is a guideline for identifying the variables of research and the hypothetical relationship between them.

### Establishing dependent, independent and intervening variables

In social science research, researchers generally deal with three types of variables: dependent variables, independent variables, and intervening variables. These variables can have a causal relationship or correlate with one another. The specification of those variables and their type of relationship provide the conceptual framework for research.

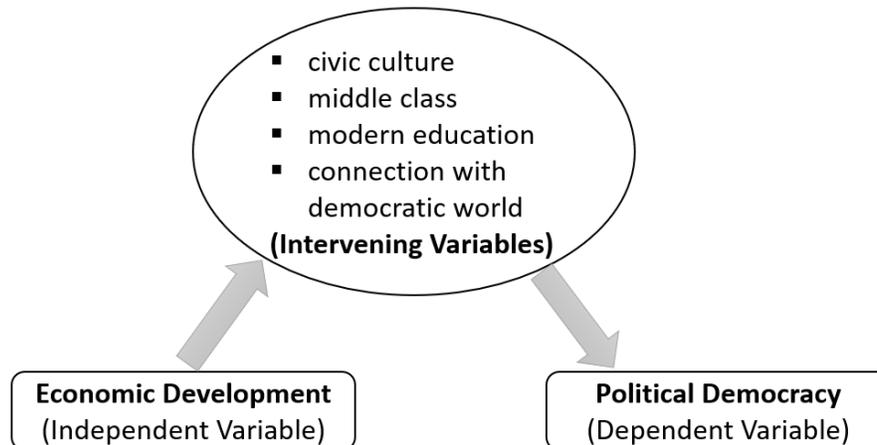
In a conceptual framework, the **dependent variable** refers to an outcome that is produced by independent variable(s). **Independent variables** are the causes of an outcome or the underlying force behind the emergence of a dependent variable. For example, if you are interested in studying democratization and searching for causes that facilitate a democratic transition, you can treat “democratization” as the dependent variable and a hypothesized causal factor such as economic development as the independent variable. See the following figure as an example.

*Figure 1: Example of independent/dependent variables*



Sometimes, independent variables do not directly lead to an outcome. Many factors intervene in the relationship between independent and dependent variables. Those factors are known as **intervening variables**. This means that, sometimes, independent factors, first produce or lead to conditions that, in turn, lead to the outcome or the dependent variable. For example, there is a generally hypothesized model that emphasizes economic development leading to political democracy as demonstrated in the above figure. This is a general model. However, there are cases where economic development does not directly lead to democratization. These instances beg the question about the role of intervening variables in the basic cause-effect relationship between independent and dependent variables or challenge the original assumption. Research, in such cases, requires the identification and inclusion of intervening variables in the causal model as illustrated in figure 4.

Figure 2: Intervening Variables example



Research topics that IWA is interested in are complex and it can be difficult to disentangle concepts of interest and the causal links between them for a given question. Miller & Salkind recommend breaking up the system you are investigating into smaller parts and considering the boundaries, interactions with other systems, and feedback loops. (Miller and Salkind 2002) Begin by asking yourself two questions: How do we achieve specific ends in the system and what key causal factors are involved? Specifically, you often want to identify a causal link between independent variables which can be manipulated or which can vary, and dependent variables that vary because the independent variables changed.

Using a theory of change can be useful for mapping out these interactions. After mapping out the key concepts, you will need to operationalize them as variables that you can measure or observe. For example, the objective of IWA's report, [Education Compromised? A Survey of Schools in 10 Provinces of Afghanistan](#), was to assess the physical condition of school buildings and the ability of the facility to deliver education services.

The two concepts of interest are *physical condition* and *ability to deliver*, which then need to be operationalized as measurable or observable variables. The researchers did this by selecting the following variables that represent these two concepts: operational status of schools, quality of school infrastructure, functionality of school support mechanisms, access to sanitation facilities, availability of trained teachers, gendered information on student enrollment, quality and quantity of learning materials, and student-teacher ratio. These variables are observable or measurable components of the two concepts of interest, which in turn are smaller components of the research objective, which was to evaluate the quality of education among donor-funded schools in Afghanistan. The process of breaking down your research into smaller and smaller sub-components is a fundamental part of any research design.

After variables are mapped onto concepts, you should decide on an appropriate research design. To make this decision, you need to consider resources (i.e. time, funding, capacity), ethics (i.e.

can I assign a treatment to one group but not another), the security situation, and existing research that you can build on.

### Internal Validity

One critical factor when deciding on the research design to use is to consider threats to internal validity, also called threats to inference, that are inherent in every research design. These threats prevent you from being able to claim that a change in an independent variable (i.e. an intervention, policy, programme, or treatment) caused some change in a dependent variable. Threats to internal validity allow for alternative explanations for the observed results. External validity, on the other hand, refers to your ability to generalize results (i.e. reproduce the study with a different sample and receive similar results).

There are eight threats to internal validity. History, maturation, instrumentation, and testing arising from single group studies. Selection bias, regression to the mean, social interaction, and attrition arise from multi-group studies.

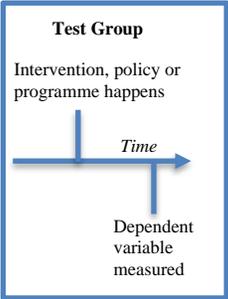
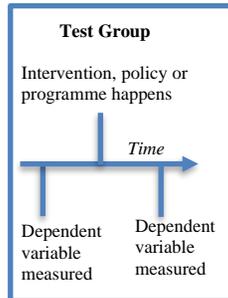
- **History:** Some unrelated event affects the dependent variable
- **Instrumentation:** The measures that are used to assess the dependent variable are different in the pre and post-test
- **Testing:** The act of measuring the dependent variable in the pre-test affects the post-test results
- **Maturation:** Natural changes that occur over time have a significant effect on the dependent variable
- **Regression to the mean:** A very low or high result in the dependent variable has a statistical tendency to get closer to the mean the second time it is measured
- **Reactivity** (also called observer effect): Interaction between the researcher and participants and among participants can cause participants to react differently to an intervention, which might cause distortions in the dependent variable
- **Attrition:** Participants might opt out, be unreachable, or pass away over the course of the study
- **Selection bias:** The test group (receives the intervention) and control group (does not receive the intervention) are not comparable at the beginning of the study and might have observed and/or unobserved differences that might impact the dependent variable

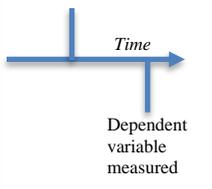
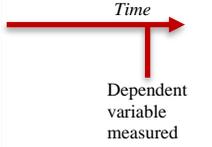
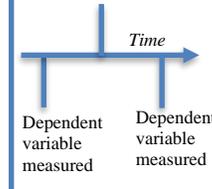
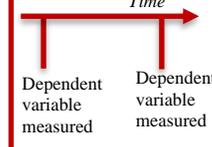
These eight factors are included in the table of research designs on the following page.

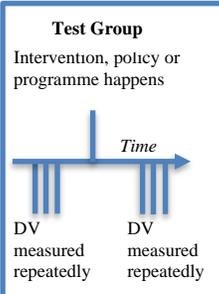
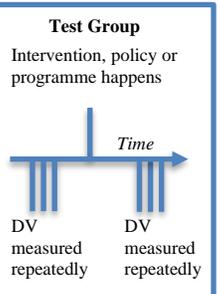
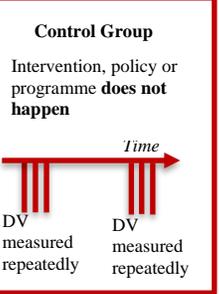
### Variable-based Experimental designs

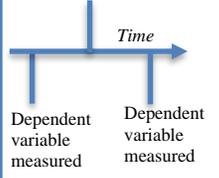
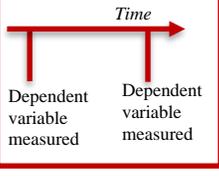
Variable-based experimental designs are useful for testing links between independent and dependent variables in order to suggest evidence of cause and effect or validate proof of concept. These designs are best used when there are objective units of measures associated with each variable. In designs where a treatment can be compared to a control group, this is useful for measuring the effectiveness of the treatment. These approaches are less effective for determining causation in complex environments.

Table 1: Variable-based research designs

	Research design	What it is	Example	When to use it	Threats to internal validity	Published example
Pre-experimental	<p><b>One-shot case study</b></p>  <p><b>No Control Group</b></p>	<p>A case study of a single group after an intervention has already happened. There is no pre-measurement before the intervention happened.</p>	<p>The government enacts a policy to ban young men from playing violent video games (intervention). One year later, you measure how many young men are committing violent crime and notice that the rate has decreased.</p> <p>This decrease, however, could have been caused by any number of factors and may have been decreasing well before the intervention.</p>	<ul style="list-style-type: none"> <li>To explore a new area of research and determine if further investigation is needed;</li> <li>Information needed quickly and for a low cost</li> <li>Security issues prevent possibility of measuring control group, over time, or it is unethical to assign treatment</li> </ul>	<ul style="list-style-type: none"> <li>Selection bias</li> <li>History</li> <li>Can't observe any change as there is no pre-measurement</li> </ul>	<p><a href="#">Education Compromised? A Survey of Schools in 10 Provinces of Afghanistan</a></p>
	<p><b>Observational study (one group pre-test post-test)</b></p>  <p><b>No Control Group</b></p>	<p>A study where a single group is studied at two points in time, once before an intervention and once after, with comparisons drawn to general expectations of what would have happened without the intervention.</p>	<p>A business measures the productivity of its employees</p> <p>The business hires a consultant to facilitate a training session for all employees. (intervention)</p> <p>The business measures the productivity of its employees afterwards and notices an increase in productivity</p> <p>This increase might be because of the intervention, but rather because of the first measurement (employees feel pressured to perform better), or because of other factors that changed during that time.</p>	<ul style="list-style-type: none"> <li>To explore a new area of research and determine if further investigation is needed</li> <li>Information needed quickly and for a low cost</li> <li>When it is unethical or impossible for there to be a control group</li> </ul>	<ul style="list-style-type: none"> <li>Selection bias</li> <li>History</li> <li>Maturation: effects due to time-related changes between first and second measurement unaccounted for</li> <li>Regression to the mean</li> <li>Possibly attrition, instrumentation</li> </ul>	<p><a href="#">Impact of a "TED-Style" presentation on potential patients' willingness to accept dental implant therapy: a one-group, pre-test post-test study</a></p>
	<p><b>Natural or cross sectional experiment without pre-test (also called static group comparison)</b></p>	<p>A study where dependent variables for two groups, one that experienced an intervention and one that did not, are measured once after the intervention. If control group is not available, can compare treatment group against aggregate benchmark data</p>	<p>One school enacts a policy requiring all students to wear a uniform</p> <p>Another school does not enact this policy.</p> <p>You measure student attendance and notice that it is higher in the school with the uniform policy than the one without it.</p> <p>While this could be due to the uniform policy, it could also be that</p>	<ul style="list-style-type: none"> <li>To determine if further investigation is needed</li> <li>Information needed quickly and for a low cost</li> </ul>	<ul style="list-style-type: none"> <li>Selection bias, no way to prove test and control group were comparable in their non-observed characteristics</li> <li>While there is a control group, which addresses history, as large-scale historical events would have affected both control</li> </ul>	<p><a href="#">Differences in students' learning outcomes between discovery learning and conventional learning models</a></p>

	<div style="border: 1px solid blue; padding: 5px; margin-bottom: 10px;"> <p style="text-align: center;"><b>Test Group</b></p> <p>Intervention, policy or programme happens</p>  </div> <div style="border: 1px solid red; padding: 5px;"> <p style="text-align: center;"><b>Control Group</b></p> <p>Intervention, policy or programme <b>does not</b> happen</p>  </div>	<p>which is usually a national or regional average for the outcome in question.</p>	<p>attendance rates were higher for that school before the uniform policy was enacted. It is also not possible to know whether the test group would have had lower attendance without the uniform policy nor can we know whether the control group would have had higher attendance with the uniform policy.</p>		<p>and test groups in similar ways</p> <ul style="list-style-type: none"> <li>• Regression to the mean</li> </ul>	
<p>Quasi-experimental</p>	<p style="text-align: center;"><b>Natural experiment with pre and post test</b></p> <div style="border: 1px solid blue; padding: 5px; margin-bottom: 10px;"> <p style="text-align: center;"><b>Test Group</b></p> <p>Intervention, policy or programme happens</p>  </div> <div style="border: 1px solid red; padding: 5px;"> <p style="text-align: center;"><b>Control Group</b></p> <p>Intervention, policy or programme <b>does not</b> happen</p>  </div>	<p>A study where the dependent variable is measured before and after an intervention for a test group and in the absence of the intervention for the control group. The test and control groups are not assigned the intervention by the researcher, but rather they are assigned by some force beyond the control of the researcher.</p>	<p>You measure school attendance in two villages in two different regions. You notice that village A has similar attendance rates to village B</p> <p>Village A in the first region experiences heavy conflict (intervention)</p> <p>Village B in the second region remains relatively peaceful (no intervention)</p> <p>You measure school attendance in the two villages and notice that attendance is now lower in village A than in village B</p> <p>While it is very plausible that drops in school attendance were due to increased conflict, some part of the difference in attendance might be due to unobserved differences between the two villages</p>	<ul style="list-style-type: none"> <li>• Impossible, costly, or unethical to assign treatment</li> <li>• Pre-intervention data of interest is available</li> <li>• When this method can be supported by approaches to create a more valid control group (see appendix B)</li> </ul>	<ul style="list-style-type: none"> <li>• Selection bias: unobserved differences between test and control group might still exist that could account for the change, if any, in the dependent variable</li> <li>• Use of control group and pre- and post-measurement of dependent variable addresses regression to the mean, effects of history, and maturation over time</li> <li>• Possibly attrition, instrumentation</li> </ul>	<p><a href="#">Does lengthening the school day increase students' academic achievement? Results from a natural experiment in Chile</a></p>

<p style="text-align: center;"><b>Time series</b></p> <div style="border: 1px solid blue; padding: 5px; margin-bottom: 10px;"> <p style="text-align: center;"><b>Test Group</b></p> <p style="text-align: center;">Intervention, policy or programme happens</p>  <p style="text-align: center;">No Control Group</p> </div>	<p>A study where multiple measurements of the dependent variable are taken before and after an intervention for a single group</p>	<p>You measure the death rate in Kabul over a number of years</p> <p>The government enacts a law banning violent video games</p> <p>You measure the crime rate in Kabul for several years afterwards and notice that it has been slowly decreasing</p> <p>While it may be possible that the law banning violent video games partially caused the crime rate to slowly decrease, there are other variables that changed during this time (i.e. inequality, poverty rate, security levels) that may have caused the crime rate to decrease rather than this policy</p>	<ul style="list-style-type: none"> <li>• When a broad policy is rolled out and it is difficult to disentangle beneficiaries from non-beneficiaries</li> <li>• Pre-intervention data of interest is available</li> <li>• Enough resources to conduct multiple measurements</li> </ul>	<ul style="list-style-type: none"> <li>• Selection bias</li> <li>• History: effects on dependent variable due to other unrelated changes from first to second measurement unaccounted for because there is no control group</li> <li>• Maturation: effects on dependent variable due to time-related changes between first and second measurement remain unaccounted for</li> <li>• Possibly attrition, instrumentation</li> </ul>	<p><a href="#">The effect of the late 2000s financial crisis on suicides in Spain: an interrupted time-series analysis</a></p>
<p style="text-align: center;"><b>Control group time series</b></p> <div style="border: 1px solid blue; padding: 5px; margin-bottom: 10px;"> <p style="text-align: center;"><b>Test Group</b></p> <p style="text-align: center;">Intervention, policy or programme happens</p>  </div> <div style="border: 1px solid red; padding: 5px;"> <p style="text-align: center;"><b>Control Group</b></p> <p style="text-align: center;">Intervention, policy or programme <b>does not happen</b></p>  </div>	<p>A study where multiple measurements of the dependent variable are taken before and after an intervention for a test group and at the same time for a control group that does not receive the intervention</p>	<p>Over a number of years, you measure the rate of healthcare utilization in two rural villages that have a health clinic within two hours of walking distance. Village B has a lower rate of healthcare utilization than village A</p> <p>A hospital is within one hour of village A (intervention)</p> <p>You measure the rate of healthcare utilization over a number of years and notice that village A has higher rate of healthcare utilization than village B</p> <p>This is likely due to the hospital (intervention), but it is also possible that other unobserved differences between village A and B caused them to change their behaviour</p>	<ul style="list-style-type: none"> <li>• Useful when a control group is readily accessible</li> <li>• Useful when treatment is assigned beyond the control of the researcher or when it is ethical and possible to non-randomly assign treatment</li> </ul>	<ul style="list-style-type: none"> <li>• Selection bias</li> <li>• Use of control group and pre- and post-measurement of dependent variable addresses regression to the mean, effects of history, and maturation over time</li> <li>• Possibly attrition, instrumentation</li> </ul>	<p><a href="#">The impact of rural health system reform on hospitalization rates in the Islamic Republic of Iran: an interrupted time series</a></p>

True experiment	<p><b>True experiment (with pre-test and post-test control group)</b></p> <div style="border: 1px solid blue; padding: 5px; margin-bottom: 10px;"> <p style="text-align: center;"><b>Test Group</b></p> <p style="text-align: center;">Intervention, policy or programme happens</p>  </div> <div style="border: 1px solid red; padding: 5px;"> <p style="text-align: center;"><b>Control Group</b></p> <p style="text-align: center;">Intervention, policy or programme <b>does not happen</b></p>  </div>	<p>A study where the intervention is randomly assigned to a test group and a control group. The dependent variable is measured before and after an intervention for a test group and at the same time for a control group that does not receive the intervention</p>	<p>You identify one hundred high school students of the same age and of similar socioeconomic backgrounds (control variables) who are interested in becoming medical students.</p> <p>You randomly assign half of them to receive a specialized training program (intervention) to get enrolled in medical school</p> <p>The other half do not receive the training program.</p> <p>You measure the number of those who successfully enrol into medical school from each group.</p>	<ul style="list-style-type: none"> <li>• Applicable when there is strong interest in a well-defined research problem with significant funding</li> <li>• Useful when it is ethical and possible to randomly assign treatment</li> </ul>	<ul style="list-style-type: none"> <li>• Unobserved differences between two groups that we did not measure are accounted for due to randomization, which addresses selection bias</li> <li>• A large enough sample size will allow us to say that the independent variable caused the change in the dependent variable</li> <li>• Use of control group and pre-and post-measurement of dependent variable addresses regression to the mean, effects of history, and maturation over time</li> <li>• While internally valid within the sample, more work might need to be done before generalizing to a broader population (external validity)</li> </ul>	<p><a href="#">Bringing Education to Afghan Girls: A Randomized Controlled Trial of Village-Based Schools</a></p>
-----------------	---	--	---	---	--	---

You will most likely not conduct any ‘true’ controlled experiments considering the security situation in Afghanistan and because true experiments outside of laboratories or tightly controlled programs are often unethical or unfeasible, take a significant amount of time and incur a high cost. Instead, you will likely be conducting pre-experimental and quasi-experimental research, where some degree of internal validity is sacrificed considering the context and sensitive subject matter. Please refer to Appendix B on approaches to increasing internal validity in quasi-experimental approaches.

Along with choosing an experimental design for understanding the interaction between variables, you also want to find ways of combining qualitative and quantitative data using **Mixed**

**Methods.** This helps gather different perspectives and analyses on different levels from institutional to social processes. The table below provides a few ways to collect the two different types of data in a complementary manner:

*Table 2: Tools for Bridging the Qualitative-Quantitative Divide*

Method	Description	Example
<b>Focus on Tipping Points</b>	Qualitative analysis can explain turning points in quantitative time-series and changes over time in causal patterns established with quantitative data.	A trend of school enrollment shows a spike in enrollment during a specific month. In surveys, you specifically ask participants to explain their reasons for enrollment during that month.
<b>Typicality of Qualitative Inferences Established by Quantitative Comparison</b>	Close qualitative analysis of a given set of cases provides leverage for causal inference and quantitative analysis then serves to establish the representativeness of these cases.	After meal programs are implemented at a school, students surveyed report that they find learning easier. A quantitative study is then used to determine if students at schools with meal programs report higher learning scores.
<b>Quantitative Data as point of Departure for Qualitative Research</b>	A quantitative data set serves as a starting point for framing a study that is primarily qualitative.	An quantitative study shows that one area has a lower maternal death rate compared to surrounding areas. Researchers recruit participants in the area to participate in focus groups where they discuss services that help improve maternal health.
<b>Sequencing of Qualitative and Quantitative Studies</b>	Across multiple research projects in a given literature, researchers move between qualitative and quantitative analysis, retesting and expanding on previous findings.	One study on food insecurity interviewed participants to find that they encounter difficulties registering for food vouchers while another study shows an increase in food voucher registration. Researchers employ different methods to study this discrepancy.
<b>Triangulation</b>	Involves confirming and corroborating results reached by one method with other results reached by another method.	Quantitative study of a company's payroll shows an increase in income. Researchers follow-up to survey participants to see if their incomes have increased.

(Source: Tarrow, 2009)

After selecting a research design, you can begin to consider [sampling](#) and note it in this section. Clarify if you will use quantitative or qualitative methods or both. For qualitative methods, refer to the section in Research and Evaluation Implementation on how to conduct [surveys](#) and [focus groups](#). Explain your data sources (i.e., primary/secondary). Detail what you will look at, the different sets of data you will gather (i.e. literature review, documents, field interviews, surveys etc.). State how this data will help you to understand aspects of your research problem and answer your questions. Explain how you will proceed with each of the methods used.

Justify your choice of methods. Whichever data collection methods you employ, make sure you demonstrate that you have put some thought into reliability and internal/external validity. You should also show that you are being realistic about what you can accomplish.

## 5. References

Append a section which includes complete references for ALL sources cited in your proposal.

Use a consistent citation and referencing style (i.e. APA, AAA, ASA, Chicago, etc.), whichever format you use, be consistent. The section on report writing at the end of the guide includes a detailed citation guide for Chicago Notes-Bibliography 17<sup>th</sup> edition.

The resources cited here will likely feature in your policy paper's references as well. Using [Zotero](#) or other reference management software will make it easy to copy and paste citations over to your policy paper, this tool is also described in more detail at the end of the guide.

Refer to the [References](#) section for more details.

## Needs Analysis

This tool is used to inform decisions by identifying **needs** which are defined as the gaps between current and desired conditions. As such, it is often described inter-changeably with gap analysis. Need analyses are usually conducted in the following circumstances:

- Planning for decision making about program implementation
- Advocacy in grant preparation or other funding requests
- Budgeting to set funding priorities
- Testimony to create community awareness, to show action on a problem or to satisfy a legislative mandate

Needs Analyses can be used at different stages of an intervention but mostly conducted proactively before an intervention or decision to identify gaps that could be addressed through potential future projects or opportunities. Sometimes, it is conducted during an intervention as an integrated component of monitoring and evaluation for a project cycle to inform the next project cycle. The benefits of needs analysis is that it provides a systematic process for providing justification and gathering different perspectives to inform decisions.

A needs analyses to guide decision makers can only be conducted if decision makers have a measurable definition of what conditions are desired or required and have a measurement of what the current condition is. A needs analysis is not supposed to find solutions but rather identify the needs for which solutions might then be proposed. Therefore, at the end, a needs analysis could be followed by a variety of different solutions, with a clear criteria on how to compare options.

### **Basic Steps**

1. Identify PARTNERS and target populations through a [stakeholder analysis](#). Conduct [Conflict Analysis](#) if necessary.
2. Define desired conditions. Look into drawing information from a [Conceptual framework](#)

3. Collect information on current conditions using appropriate quantitative or [qualitative methods](#)
4. Define needs on the basis of gaps between current and desired conditions.
5. Analyze Needs and Determine Next steps as outlined in the [Policy Brief section](#).
  - Establish prioritization of needs on the basis of costs associated with meeting the needs vs the cost associated with not meeting the needs. This can be done using a [cost-effectiveness analysis](#).
  - Assess what works, what doesn't, and the systemic relationship among needs.
  - If possible, collect information on causal factors, leading to priority needs.

## Evaluation Design

### Overview

Evaluations reveal what is being done right, learning opportunities, impacts, and side effects of interventions. They describe the relevance, efficiency, effectiveness, impact, and sustainability of an intervention. They can confirm if your causal assumptions about a problem or issue is being addressed. In this section, the terms intervention and programs are used interchangeably since evaluations can be done for both. However, while programs usually describe specific activities, interventions can globally include policies and programs implemented to effect change. Evaluations are especially useful when there are unexpected results or performance outliers, when there is a long period with no improvement without a clear explanation as to why, or when similar programs or policies report divergent outcomes or when indicators for the same outcome show divergent trends. They are also helpful when a project is launching a new approach such as a pilot. Unlike need analyses, program evaluations focus on what IS being done rather than what SHOULD be done.

On the one hand, most evaluations for programs are conducted within the **Result-Based Monitoring and Evaluation (RBME)** framework which compares what was planned with what was actually delivered and demonstrates effectiveness of intervention in relation to objectives. It is usually conducted once a project is sufficiently developed. Result-based evaluations that answer descriptive and normative research questions provide a summary statement to donors to demonstrate accountability and document program activities in the greatest details.

On the other hand, evaluations that examine impact provide advice to program developers on how to improve the program by delving into trends or problems to describe impact mechanisms and causality. **Impact evaluations** are used for anticipating and measuring the impact as well as the outcomes of an engagement. They can also contribute to a global evidence base of what works for whom and in what situations. The OECD-DAC defines impact evaluation as the following:

“positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.”

An effect can be understood as a change in social, environmental or economic outcome that is more long-term. An impact evaluation also emphasizes looking at negative or unintended effects that are not laid out in program plans. Furthermore, an impact evaluation helps inform decisions about continuation, expansion, or replication of programs.

The two types of evaluations are complementary rather than substitutes. Impact evaluations are often more resource-intensive. They go further in trying to determine whether changes are due to intervention, whether it is possible to identify program effect, and if effect can be attributed to program instead of other causes.

This section first describes the pre-assessments necessary for evaluations: Conflict Analysis, Stakeholder Analysis, Theory of Change, and Evaluability Assessment. The first three can be used universally for all studies including research while the latter two are specific to evaluations. The section then presents different types of question including result-based management and process questions that evaluations seek to address. Each type of question also includes recommendations on which of the three approaches to take for designing an evaluation: variable-based, process-based, or participatory. The final covers Peace and Conflict Impact Assessment, a unique form of evaluation that focuses on interventions with peacebuilding aspects.

## Pre-Assessment Analyses

Conflict Analysis, Stakeholder Analysis Evaluability Assessment, and Theory of Change help understand the context in which an intervention is implemented. Pre-assessments also make the evaluation easier by establishing entry points through the stakeholder analysis and a shared understanding through a theory of change.

- Conflict Analysis (*context-specific*)
- Stakeholder analysis
- Theory of Change
- Evaluability Assessment

These pre-evaluation assessments investigate an issue or problem in order to determine what is known about it and how to proceed in developing an evaluative approach to it. They provide the context needed to ask the right questions. The conflict analysis assessment is not applicable for all contexts and is most useful when you are analyzing interventions in fragile and conflict affected states.

### Conflict Analysis

A conflict analysis, also sometimes called a strategic conflict assessment, looks at the conflict profile and dynamics of a particular conflict. It helps you gain a better understanding of the context that you are working in. It presents a static description of the conflict profile that can be used to identify different dynamic scenarios. Information for producing the conflict profile and dynamics can be collected through desk research, surveys, expert interviews, community consultations, or pre-existing knowledge.

To create a **conflict analysis profile**, there are generally three types of causes:

- **Structural:** pervasive fundamental factors that have become built into the policies, structures and fabric of a society and may create the pre-conditions for violent conflict
- **Proximate:** factors in recent history contributing to a climate conducive to violent conflict or its further escalation, sometimes apparently symptomatic of a deeper problem
- **Triggers:** single key changes, events or their anticipation or exacerbate proximate causes and lead to the (re-) emergence or prolonging of conflict? How likely are they? Understand the conditions that could lead to these triggers.

This template for conflict analysis profile helps you to identify the most important sectors to look for conflict causes. The table below provides some questions that you can look into answering for each sector. You do not have to answer all the questions from each sector- just the relevant ones.

Note that this framework often provides a higher-level analysis than used at a national level and may not be as suitable for more local level analysis. Although not implied in the template, the profile analysis should look for **dynamic interaction between sectors** as well.

*Table 3: Conflict Analysis Template Drawn from the Country Indicators for Foreign Policy (CIFP) Fragile States Methodology*

	Structural causes or mitigators of conflict	Proximate causes or mitigators of conflict	Key risks, potential “stresses or triggers” and how likely they are	Current status	Overall trajectory and trend
Political, social economic elites	Who are the key elite groups and what is their power basis? How do they bargain with each other? How credible are their agreements; is there an 'elite pact' / 'political settlement'? How are rents, power, resources etc. distributed?	Culture of fear and mistrust	I.e. elections, reform processes, arrest / assassination of key leader or political figure, military coup	How severe are these cleavages?	How do you expect these cleavages to evolve over time?
Political, social economic groups	What/ who are the key social groups (i.e. ethnic religious etc.)? What is the relationship between them/ how do they relate to each other? Are some groups systematically excluded/ marginalized from political, social or economic participation??	Disruption of social networks, culture of fear and mistrust, lack of political participation, human rights abuses	I.e. elections, radicalisation of conflict parties,  Mitigators: communication channels between opposing parties, civil society commitment to peace, anti-	How severe are social cleavages?	How do you expect these cleavages to evolve over time ?

			discrimination policies.		
Political system	How robust is the political system? (How) Is it institutionalized through the political system? (How) does the political system mitigate or reinforce/ amplify elite and social divisions?	Illegitimate governance, decentralisation		How robust is the political settlement process	
“Social contract”	What is the relationship between citizens and the state? Do citizens perceive their government as legitimate? (How) can citizens articulate their expectations and (how) responsive is the state to these expectations? How institutionalized is this relationship through the political system?	Culture of fear and mistrust	reform processes, arrest / assassination of key leader or political figure,	How robust are state society relations	How do you expect citizen-state relations to evolve over time?
Security	Development of a war economy	Military and civilian deaths, presence of armed forces, uncontrolled security sector, light weapons proliferation	Establishment of paramilitaries,  Mitigators: demobilisation process, reform programmes,		
Economic development	economic structures, opportunities available to different groups, access to natural resources		Sudden collapse of local currency, increased price/scarcity of basic commodity, capital flight		
Human development (service delivery i.e. health, education)		Prevalence of disease and health issues, malnutrition, poverty,	Rapid changes in unemployment, increased price/scarcity of basic commodity		
Demographic and population make-up	Age distribution of population (i.e. youth bulge)	Migration, return of refugees or internally displaced persons (IDPs)			
Geography and environment	Physical geography and geo-strategic position, frontlines around the location of natural resources	Destabilising role of neighbouring country, climate change impacts	Natural disasters		
Conclusion		Based on the above, what are the key issues	What are the most likely key risks/ potential stresses?	How severe is conflict ?	How do you expect this conflict to evolve over time?

		arising from your consideration of the nature and causes of conflict?			
--	--	---	--	--	--

This template should be used to guide a conflict-sensitive evaluation instead of just creating a general conflict mindfulness. You need to clearly describe how different contexts produce different outcomes. Although conflict is often understood as violent disruptions, you should not overlook broader state-society conflicts, localized conflicts and other dynamics, like inter-religious, gender-based or intergenerational conflict. This tool is good at pointing out horizontal dimensions of conflict such as like historical marginalization by central government, as well as vertical ones like conflict between communities.

Drawing from the overall trajectory and trends in the profile, create scenarios to highlight conflict **dynamics**. Creating these scenarios helps you anticipate different kinds of engagements that you may need to take. **Scenario** should also provide information regarding how different challenges and opportunities may be realized or overcome.

*Table 4: Conflict Analysis Scenarios template*

Time horizon	Determines the degree of predictability and the types of elements most likely to prove decisive	
Scenario types	Structure + Trends + Stakeholder behaviour	Window of opportunity for engagement
Best case		
Worst case		
Most-likely		
Wild-card		

For examples on how to create conflict analysis and fragile states profiles and dynamics, take a look at the [Country Indicators for Foreign Policy Reports on Fragile and Conflict Affected States](#).

#### Stakeholder analysis

Stakeholder analysis helps identify those who influence the intervention and anticipates the kinds of influence they may have on the evaluation. Conducting this analysis helps you develop strategies to get the most effective support possible or buy-in for the evaluation and reduce obstacles.

## Steps to Conducting a Stakeholder Analysis

1. Brainstorm to identify people, groups, and institutions that will affect or be affected by the intervention. Stakeholders can also be drawn from the conflict profile's elites and groups. List them in the stakeholder column of the table. Although not indicated in the table, highlight the important relationships between actors at various levels, and their perception of these relationships.
2. Identify the stakeholder type: **Primary stakeholders** are directly and immediately affected by or involved in the intervention. They include the **main client** who requests the evaluation and who you are accountable to. Other primary stakeholders include those with the power (however defined) to initiate developments within a situation and **beneficiaries** who are most affected by the intervention. **Secondary stakeholders** are indirectly impacted or impacted in the aftermath and tend to respond to other actors' actions rather than 'drive' events.
3. Identify the specific **interests** each stakeholder has in the intervention. Interests are the underlying socio-political agenda or motivations of the actors (concerns, goals, hopes and fears). Record these interests in the column labeled "stakeholder interests." These interests also help determine the potential benefits or costs of the intervention or evaluation to the stakeholder. For example, community leaders who may benefit disproportionately from projects may be adverse to evaluations that suggest more equitable distributions of power.
4. Based on each stakeholder's interests, determine how their potential **actions could affect the success of the intervention or evaluation**. Consider both (a) the role the key stakeholder must play for the evaluation to be successful and the likelihood that the stakeholder will play this role and (b) the likelihood and impact of a stakeholder's negative response to the evaluation. Record each stakeholder's assessment under the column labeled "assessment of potential impact" by assigning an "A" for extremely important, a "B" for fairly important, and a "C" for not very important.
5. Consider the kinds of **actions you could take to gain stakeholder support and reduce opposition** for the evaluation. Consider how to approach each of the stakeholders. What kind of issues will the stakeholder want the evaluation to address? How important is it to involve the stakeholder in the planning process? These are critical considerations for the success, and especially for the relevance of the project - especially the involvement of the less powerful and marginalized participants in the process (often not heard because those who have the power in the situation assume, they know what is most needed).
6. Some only need to be informed of the process and others should be involved in decision-making capacities depending on the resources and relationships they offer. The greater participation of some stakeholders could imply greater costs. Are there other groups or individuals that may influence the stakeholder to support the evaluation? Record the

strategies for obtaining support or reducing obstacles to the evaluation in the last column of the table.

Table 5: Profiling stakeholders

Stakeholder	Type of stakeholder	Stakeholder interests	Potential impact of evaluation on stakeholder	Potential impact of stakeholder on evaluation	Potential strategies for obtaining support or reducing obstacles

(Source: Management Sciences for Health and the United Nations Children’s Fund, 1998)

### Theory of Change (TOC)

TOC conveys beliefs about the steps that are necessary for an intervention to succeed in reaching its objectives and helps you confirm whether an intervention’s results are appropriate for addressing the problem. The TOC puts forward a definition for impact which could be short or long-term. Impacts that are longer-term are often less specific and more abstract concepts such as governance, conflict reduction, and cultural change.

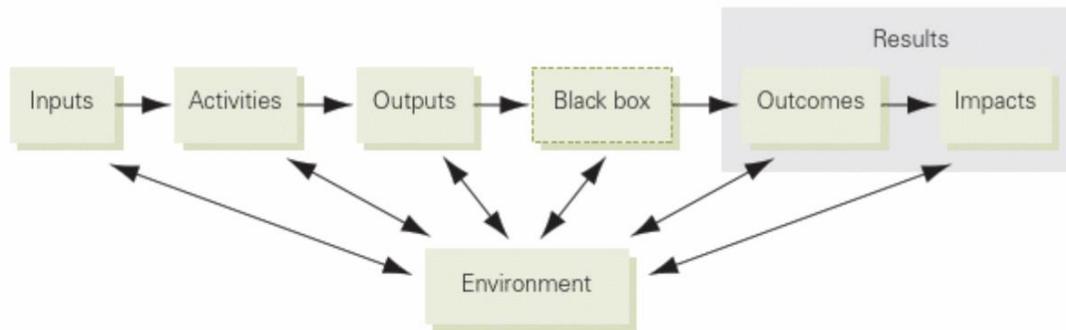
Your TOC should be **supported by research and evaluation underlying the program** that comes from the knowledge fund and from the local/contextual knowledge and experience of participants. If a TOC already exists before an evaluation, it should be reviewed using the [checklist for TOC components](#) (appendix A).

To start, identify the **problem** and the **community’s needs** (see results from needs assessment) in the current situation. Then identify the outputs, outcomes and impacts in the situation that you want to achieve through the intervention. **Outputs** are the specific program (i.e., activities, services, products, goods) or policy that an intervention budget. **Outcomes** are short-term and intermediate changes or benefits resulting from the outputs while **Impacts** are longer-term. Note the expected trajectories of impacts as well because change does not always happen linearly.

Next, identify events or conditions that may affect the outcomes. This involves the broader context or **influential factors** such as political environment, public attitude, aid players, macroeconomic picture, or policy context that influences all parts of the system. These factors can happen before or during the program’s progress. These can be drawn from the stakeholder or conflict analysis.

Finally, fill in the inputs and activities that are necessary to get you from the current to the desired situation. **Inputs** are tangible resources (i.e. financial, material, and human resources) and **activities** use inputs to create the outputs.

Figure 3: Theory of Change components

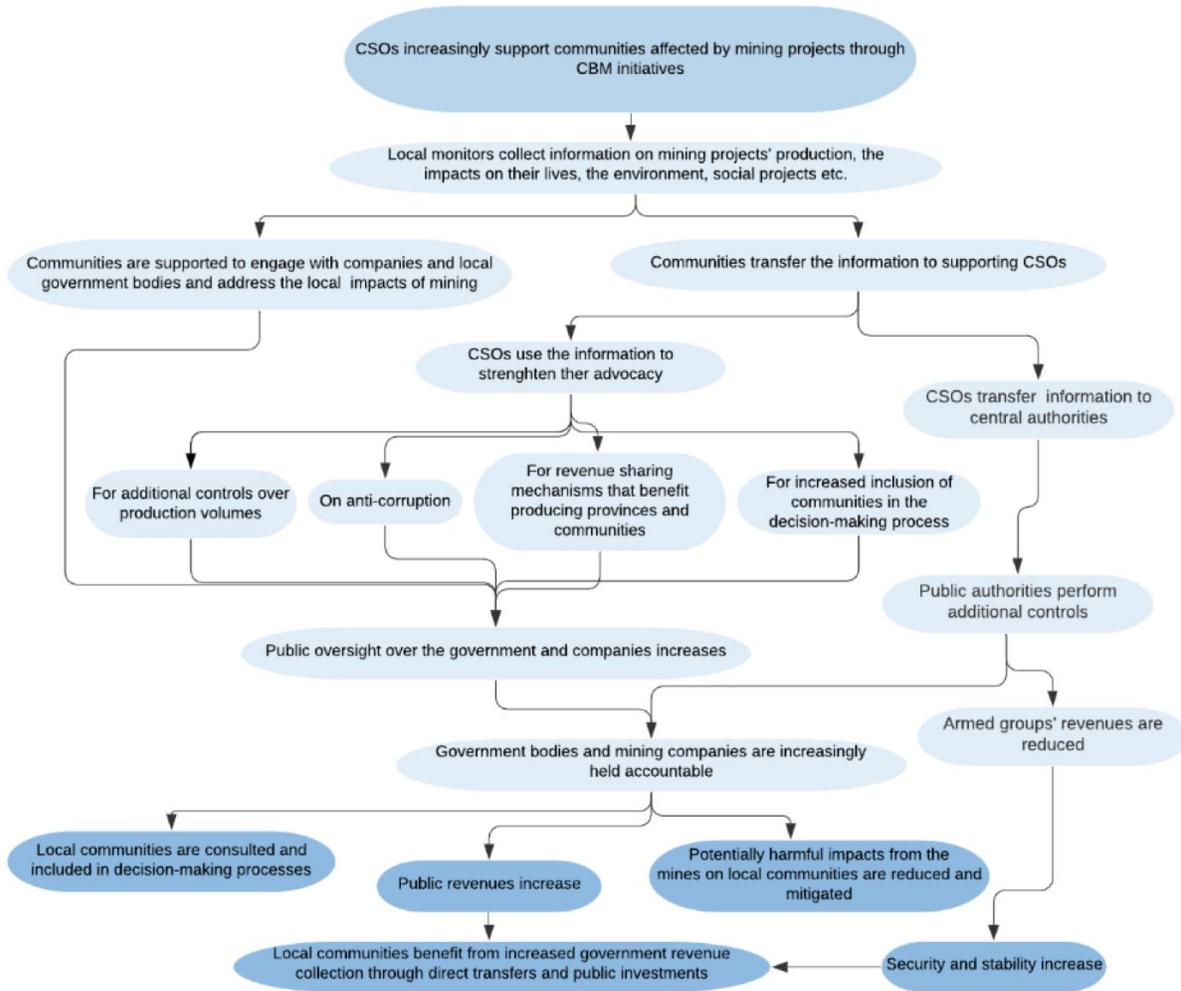


(Source: Imas and Rist, 2009)

Finally, identify critical **assumptions** that, based on the policy and environmental context, explain the connections among activities, outcomes and impacts. These assumptions also explain how influential factors will likely support or hinder progress toward results and the extent to which results are attained. For more tips on how to create a Theory of Change, refer to the [checklist in Appendix A](#).

These elements could be drawn out in a logical model that portrays a sequential order from inputs to impacts describing the assumptions on causes and effects. To create this model, work backwards from long-term goals and hypothesize if-then statements to link causes to effects. Sometimes the model might follow complex instead of linear relationships. It is important to verify a logical model with key stakeholders. A weak TOC tends to express only the intentions of policy makers and ignores actions and intentions of other stakeholders.

Figure 4: Theory of Change example



(Source Integrity Watch Afghanistan, 2020)

### Evaluability Assessment

An Evaluability Assessment is a process that helps you to identify whether and how evaluations might be useful and feasible by comparing stakeholders' expectations of the evaluation with the reality of program activities. It also defines organizational capacity and barriers and environmental influences to determine the costs of conducting such an evaluation. Evaluability depends on access to collected data, evaluation relevance and conditions, and shared evaluation perspectives among program implementers.

For an intervention to be evaluable, at the very minimum, there should be **relevant performance data or capacity to collect data** for evaluation. This means access to quality **baseline data** or targets to monitor data against. Usually, you gather this from secondary sources such as well-

documented progress reports, field visit reports, reviews and previous evaluations, and information from individuals, general public, or trained observers. If no baseline data is available, you can consider conducting quasi-experimental quantitative analyses outlined in Appendix B. To assess capacity, you should consider the data collection instruments you have at your disposal and your ability to procure data on a regular, cost-effective, and timely basis. When calculating the costs of data collection and analysis methods, include the likely monetary, political and bureaucratic costs. To create a likely timeline, plan backwards by looking at scheduled dates for an evaluation and estimated time needed to prepare and conduct evaluation to anticipate when the process needs to begin. Present examples of the types of information that would be produced explain the implications of the status quo option (no further evaluation) and the costs and potential uses of various other evaluation options.

You should also assess how **relevant** the evaluation would be given the evolving context. For example, when planning an evaluation in crisis settings, increasing tensions between different parties may impede evaluation efforts. **Political, social and economic factors** may also affect the effective conduct and use of the evaluation.

Finally, you should meet with program implementers from managers to operating-level staff to see if they share the same motives behind the evaluation and that their understanding of the program reality is similar to program expectations. Program implementers should be able to identify and agree on key resources and theory of change as well as data and evidence that have been gathered throughout the program's duration. When gathering perspectives from different levels, make sure that the views of the more powerful don't receive disproportionate weightings. Information gathered from these meetings should be compared with program-authorizing legislation or other contract documents outlining the initial program expectations.

Evaluability assessments do not only answer if evaluations are possible, they also help identify how to best adapt them to the current situation. Do not be afraid of modifying original evaluation plans at this stage. It is better to not have a fixed belief when scoping out an evaluation.

## Results Based Monitoring and Evaluation Design: Descriptive and Normative Questions

Evaluation designs depend on the types of questions the research seeks to answer. Program evaluations generally fall under two types of questions: Descriptive and Normative questions. These questions can generally be drawn from the inputs, outputs, and outcomes of a theory of change.

***Descriptive questions:*** seek to determine what the current situation is. For example, answers to descriptive questions could describe the political connections and beneficial ownership of mining companies, assess the quality of school infrastructure, or gather the perceptions of positive and negative externalities of extractive industries on community lives.

**Normative questions** compare “what is” to “what should be.” For example, answers to normative questions can help you assess if hospitals have an adequate level of equipment, medicine, and pharmacies or follow infrastructure standards.

For both of these kinds of questions, it is useful to follow a results-based monitoring and evaluation (RBM&E) system. **RBM&E** involves collecting, analyzing, and evaluating information on key indicators in order to measure progress toward clearly stated goals. Integral to this system are outcomes, indicators, and targets.

**Outcomes** are short-term and intermediate changes or benefits resulting from the outputs. They are often framed as the objective of program research since they focus and drive evaluation resource allocation and activities. Relevant outcomes to measure are those with the highest levels of resource investment or ones that are associated with anticipated problems. Since outcomes frame the research focus, they need buy-in from all key stakeholders through political process. Outcomes should also have available baseline data and appropriate available data collection methods.

*Example: the capacity of a school facility to deliver education services*

**Indicators** are variables that allow for the verification of changes through a program or show results relative to what was planned. They drive data collection and analysis. Choose indicators based on the theoretical modeling (i.e. theory of change). Indicators should be clear, relevant, adequate, monitorable in an affordable way. The number of indicators depend on how many are needed to answer the question "has the outcome been achieved?" Normally there are between 2-7 indicators per outcome. All indicators should be stated neutrally, not as “increase in” or “decrease in.”

*Example: student-teacher ratio per province collected through school surveys*

**Targets** are quantifiable levels of indicators you want to achieve by a given time. They are especially important for normative evaluations that compare reality with what is expected. Generally, if the theoretical logic of the TOC is sound, a program that achieve its targets will achieve its intended outcome as well. There is usually one target per indicator. Targets can also be set as a range or as a disaggregated set of targets for different subgroups or for different timeframes. Measuring progress towards a target usually requires a clear understanding of the **baseline** and average data over the recent past. **Baseline data** is the first measurement of indicators to find out the starting point of an intervention, as covered in the previous section on evaluability assessment. As mentioned in the Theory of Change, it is also important to consider influential factors such as the level of inputs, political concerns or organizational capacity that affects an organization’s efforts towards reaching a target.

*Example: national target of 35:1 student-teacher ratio*

To measure performance on different indicators, these types of evaluation questions are best assessed using [variable-based approaches](#) mentioned earlier under research design. In Quasi-experimental and experimental research designs, an activity of a program can serve as an independent variable and the output or outcome can serve as the dependent variable. This approach can also be used to compare outcomes before and after interventions take place.

Implementing [participatory approaches](#) could also be important for involving necessary stakeholders in selecting outcomes, indicators, and targets to make it relevant to local knowledge.

### **Note on measuring vulnerability**

Rather than focus on the symptoms of vulnerability, vulnerability analysis should focus on how an intervention impacts the **structures or sources** of inequity and inequality. For example, a program that increases food and expenditure spending for vulnerable women and children does not improve their economic negotiating powers. Deep-rooted inequities are still not affected or dismantled.

Analysis should also note instances where intervention may reinforce or exacerbate inequalities and injustice that underpin vulnerability. While most interventions have some focus on vulnerable groups, empowering one group may reinforce discriminatory practices for other groups. Increasing rights in one area might indirectly lead to negative effects on other rights. For example, increasing household income might encourage families to keep their children working at home rather than going to school which inadvertently causes an increase in child labor.

It is often harder to assess political rather than technical changes if vulnerabilities actually decreased or if inequities were addressed. Using a vulnerability lens in evaluation requires looking beyond the outcomes and placing results in context. Listed here are some examples of questions and indicators that examine vulnerability.

### **Questions that examine vulnerability and power imbalances**

- Does the intervention address inequities or the implicit violence that divides societies?
- What real control does the person have after the intervention is implemented?
- Does it assess conditions of vulnerability such as socio-cultural, poverty, lack of education and gender? Is the change enough?
- Does it assess relationships and community dynamics focused on an empowering disempowering continuum?
- Does it assess aspirations and the ability to cope? Does the intervention promote resilience?

### **Indicators for vulnerability**

- Vulnerability benchmark: This baseline for vulnerability gives a clear or comparable standard or point of reference which would provide a baseline for determining project impacts i.e., poverty line. However, recognize that vulnerability is evolving, not static and

could have multiple definitions. Therefore, it is hard to give an objective measure for a subjective concept such as powerlessness. You need to assess emerging and potential risks to vulnerability as well.

- Self-empowerment indicators: These include increased capacity to cope with vulnerability and to exercise greater control over one's own life
- Relational empowering-disempowering continuum: These include interactions that affect empowerment in a pull and push direction

## Impact Evaluation Questions

While descriptive and normative questions in program evaluations often focus on what *is* accomplished within the scope of the projects' targets and indicators, it is also important to understand the *how* and *why* these results are attained in order to assess their overall impact.

Four kinds of questions are integral to Impact Evaluations:

1. To what extent can a specific (net) impact be attributed to the intervention?
2. Did the intervention make a difference?
3. How has the intervention made a difference?
4. Will the intervention work elsewhere?

This section will suggest the most appropriate approaches towards answering each question. For the second to fourth questions, you will also be introduced to process-based and participatory approaches that are more appropriate for addressing how and why questions.

### **Question 1: To what extent can a specific (net) impact be *attributed* to the intervention?**

This kind of question generally asks if the intervention caused the intended impact. To do this, you can compare actual and expected results and processes between the intervention and desired outcome outlined in the theory of change. Remember that you are interested in the particular intervention rather than a generalization of other influential factors.

When answering questions about attribution, it is important to address the **counterfactual** - what would happen if the intervention did not happen - sometimes represented by a control group. [Quantitative experimental or quasi-experimental variable-based](#) approaches are good for making the comparison between control (no intervention) and treatment (with intervention) groups. In cases where the counterfactual is hard to establish, refer to [Appendix C on alternatives to establishing counterfactual for drawing causation](#).

### **Question 2: Did the intervention make a difference?**

While the previous question focuses on attribution, this question asks about the intervention's contribution. **Attribution** makes a claim about an intervention as a cause of impact. **Contribution** on the other hand, makes a causal claim about whether and how an intervention *among many other factors* has contributed to an observed impact. The difference is that contribution recognizes that there may be other intervening variables exterior to the intervention that lead to

the impact while attribution does not. This question addresses if the intervention was necessary or sufficient to create the impact or if other mitigating or external factors lead to the impact. In doing so, it attempts to address if these impacts would have happened without the intervention.

This kind of question can be answered with the same [quantitative quasi-experimental approaches](#) as the previous question. However, this question is usually harder to answer solely with statistical and quasi-experimental methods because of the need to consider multiple influential factors. Therefore, it draws more from process-based approaches including contribution analysis and process tracing.

### **Question 3: How has the intervention made a difference?**

This question focuses on the mechanisms and causal factors behind the intended and unintended impacts of the intervention. It also seeks to examine for whom the intervention has made a difference. This question can best be answered through flexible qualitative and mixed methods evaluation designs that are better used for describing processes. Often this question is used to assess more complex contexts with an extensive range and scope of activities that cut across sectors, themes, and geographic areas. This is why process-based approaches are best used towards addressing this question.

### **Question 4: Will the intervention work elsewhere?**

This question asks if a single case study or 'pilot' can be transferred elsewhere and scaled up or if its results are sustainable over a longer period of time. It looks for generalizable lessons and values in one setting that can be applied and implemented elsewhere on a wider scale. This involves determining a number of success factors such as the faithful application of protocols and the advantages that a particular setting presents that are needed to replicate an intervention's success. To answer this question, you need a good understanding of the contexts that created the conditions for impact. Process-based approaches are best used towards addressing this question as well.

## Addressing Impact-Related Questions Through Process-Based Approaches:

Process-based approaches are tailored towards answering broader level how and why questions. These approaches are optimal towards establishing construct and external validity:

- **Construct validity** is the degree to which you have measured what you intend to measure. Higher construct validity means that your evaluation is actually measuring the factors that are most important towards a project's success. This is because these approaches can assess for unintended and sustained impacts.
- Projects with high **external validity** can have their findings generalized and extended to other settings. As such process-based approaches are good for investigating contributions to higher level impacts as well as sustained impacts.

The weakness of process based approaches is that they are more prone to bias due to less rigorous data collection methods. There could be bias in selecting participants and bias in correlating different events.

**Contribution Analysis** assesses the contribution of other factors and helps to create a ‘contribution story’. It creates a ‘contribution story’ through identification of and evidence gathering about the logic of the links in the theory of change. In this way, it strives to see if the chain of results assumed by the theory of change occurred or not.

You can also gather evidence from qualitative and quantitative data collection methods to test against the theory of change to consider how significant the intervention is in the context of other influencing factors. This includes evidence focused on results (outputs and outcomes) and other influencing factors that can be used to test the links, assumptions, and risks in the TOC.

**Process Tracing** is similar to Contribution Analysis in that it compares results to the theory of change. However, instead of collecting a larger picture on different contributing factors, it focuses on determining the strength of causal inference based on the intervention’s ability to fulfill necessary or sufficient conditions described in the table below. Rarely are interventions able to fulfill both necessary and sufficient conditions. However, evaluators can better understand the intervention’s efficiency

*Table 6: Sufficient vs Necessary Conditions*

	<b>Sufficient:</b> Intervention would bring about the impact whenever and wherever it occurs- must be generalizable	<b>Not sufficient</b>
<b>Necessary:</b> Only way for the impact to occur is with the intervention- harder to establish	The intervention always leads to the intended effect and is the only way to get there.	The intervention is a necessary precondition for intended effects but won’t make them happen without other ‘helping factors’.
<b>Not necessary:</b>	The intervention is one way to arrive at the effect but there are other ways.	

(Source IWA, 2020)

This analysis is usually conducted during implementation rather than afterwards as an early indicator of program effectiveness. By examining the consequences of program outputs, this process-oriented approach views impact as a conjunction of results that follow a sequence of causal links instigated by the intervention.

**Case studies** focus on a particular unit and are especially useful in complex environments and for making or adjusting policy or practices to novel or unique interventions. They often use a combination of quantitative and qualitative data. Case studies interpret and describe situations to understand how different elements fit together to produce observed impacts. For example, for a transportation program evaluation, instead of gathering information on a national scale, it may be more manageable to gather data with a more narrowly defined geographic area or in several randomly or purposively selected areas. It is usually easier to understand the mechanism of a program at a smaller scale through a case study.

## Participatory Approaches

This approach establishes factors that contribute towards change by talking to stakeholders using semi-structured approaches rather than survey instruments. It ensures that beneficiaries have a voice as active agents in order to improve program targeting and relevance. Stakeholders are involved in selecting outcomes, indicators, and targets and their perspectives are included in lessons learned and impact evaluation conclusions. This helps foster ownership and build group decision-making to ensure there is a common understanding of uses that would be made with the resulting information. However, consensus should not be forced, and alternative viewpoints should be considered and recorded throughout the planning process. More than just a way to include project participants, this method also has strong external validity and is good at tackling complexity. The three evaluation designs for this approach include outcome harvesting, most significant change, and participatory impact assessment.

**Outcome harvesting** is centered on collecting and analyzing evidence as a team on what has been done differently and determining how far impacts have been created due to an intervention. The approach is different because it doesn't start with predetermined outcomes, and measure progress towards them, but rather, collects evidence of what has been achieved in the program or project area, and works backwards to determine whether and how the project or intervention contributed to the change. This is helpful especially in situations where it was not possible to gather baseline data before a change or an intervention occurred.

It is flexible and straightforward enough to be used in complex, rapidly-changing contexts. Moreover, it empowers front-line staff, communities and partners to 'monitor' what matters to them and quickly spotting changes. Staff and not just evaluators are challenged to think beyond evidence and evaluations towards impacts.

To successfully harvest outcomes, keep records of who has done what differently. Then hold review conversations regularly with your entire team to note down significant changes linked to conflict and context analyses. Workshops, using collected information can discuss the meaning and significance of outcomes in relation to context. Workshops allow teams to rank outcomes according to levels of significance. The ranked outcomes can then be shared with everyone, including donors online.

**Most significant change** collects and selects stakeholder accounts of significant changes. It respects stakeholders' judgements but validates stories through triangulation. You can collect responses from those directly involved with an intervention in a less structured way. Open-ended questions help gather participants' personal experiences.

**Participatory Impact Assessment** focuses on identifying impact for participants and the role of a program in achieving those impacts through community-based forums. Usually through focus groups, participants are asked to categorize (program and non-program) and rank factors that led to positive and negative changes. Participants also have a role in selecting indicators and methods, analyzing and triangulating information, and verifying results with the community. The

benefit of this design is that there is more “buy-in” from the community and it is more focused at the local level rather than a broader one. Parallels can be drawn between this kind of assessment and the Case Study design identified earlier.

## Peace and Conflict Impact Assessment (PCIA)

### Overview

There are two paths for integrating conflict sensitivity into evaluations: Analyzing interventions that have peacebuilding aspects and integrating conflict sensitive practices into evaluation approaches. This section will focus on the first path better known as Peace and Conflict Impact Assessments while the second path is integrated at various points in the ethical considerations and evaluation design sections.

Peace and Conflict Impact Assessment is [defined](#) as “A means of evaluating ... and anticipating the impacts of proposed and completed development projects on:

- 1) those structures and processes which strengthen the prospects for peaceful coexistence and decrease the likelihood of the outbreak, reoccurrence, or continuation of violent conflict, and;
- 2) those structure and processes that increase the likelihood that conflict will not be dealt with through violent means” (Hoffman, 2004).

Peace and Conflict Impact Assessments are a type of impact evaluations that require more consideration towards how the evaluation is designed to tailor it to conflict situations. Therefore, this section will go through the major sections of evaluations: pre-assessment analyses, questions and approaches while highlighting specific things to take note of when evaluating interventions with peacebuilding intentions.

### Key Words

Conflict Impact assessment, conflict sensitivity, do-no harm principles, Conflict/context indicators, Interaction indicators

### Pre-assessment analyses

- **Conflict Analyses:** Conflict Analyses are especially important for PCIA's because they help understand the wider effects of the intervention on the environment it operates within. This will be useful later on to help identify key indicators and issues that need to be measured to understand the peacebuilding implications of the intervention. The conflict analysis should be relevant to the regional level that program you are interested in investigating while also establishing links between different regional levels.

Figure 5: Interrelated Levels of conflict analysis



(Source: Chapter 2: Conflict Analysis, 2015)

- **Stakeholder Analyses:** When operating within a polarized environment, the interests and actions of different actors are hardly ever politically neutral. Recognize that there are certain actors who may be spoilers. Local ownership can unintentionally come to mean ownership by conflict parties, or by the most powerful sectors of society. Informants or stakeholders may use disinformation as a tactic. Even for donors and project staff who may attest to political neutrality, giving aid in conflict settings cannot and does not have a neutral impact on conflicts where it is provided. How resources are organized and delivered, play into and reinforce relationship between contending groups in recipient societies. Stakeholder analysis will also help you plan out how you want to secure political buy-in while ensuring impartiality. Building trust with key stakeholders in PCIA is crucial since you are working in politically sensitive environments. It is common for project staff and participants to be skeptical towards evaluations that could reveal power dynamics or unsatisfactory project results.
- **Theory of Change:** A TOC can provide a causal story of an intervention. It is used as a benchmark for testing assumptions. Assumptions that underpin a TOC do not just include stories of how an intervention might work but also include the project's approach towards conflict resolution; the way practitioners define their roles, values, and mandates in conflict situations; and how far they are able and willing to implement a coherent and effective set of measures to establish conflict-sensitive policies within the organization. Since peace and conflict often have complex causal factors, a TOC is useful for understanding other influential or intervening factors that might affect a project's results. Both a TOC and Conflict Analysis provide an outside-in analysis which starts from a broader understanding of the conflict and environment that the project is located within.
- **Evaluability Assessments:** For PCIA, consider project capacities and project-environment correspondences. Capacities include having the right resources, organization experience, policies, institutional structures, and tolerance to change. Project-environment

correspondences refer to the level of political support from different levels of authority and community buy-in.

### PCIA Evaluation Questions and Approaches

While all three approaches: variable-based, process-based and participatory approaches are relevant towards PCIA, process-based and participatory approaches are the most relevant for understanding higher-level complex processes towards peacebuilding. This section will go through all three approaches by highlighting the strengths and weaknesses of each towards PCIA.

#### **Variable-based approaches for PCIA**

Using a results-based framework of measuring indicators and outcomes does not always present the easiest and most appropriate way of measuring how well peacebuilding interventions achieve their objectives. This is because rather than producing tangible outputs such as hospitals or increased income, peacebuilding initiatives work towards more processes-oriented outcomes such as improved behaviors that can take on different subjective interpretations. Furthermore, it may be impossible to impose a quasi-experimental research designs of comparing against a control case to establish a counterfactual.

Nevertheless, when selecting measurable indicators, it is suggested that you choose a kaleidoscope or variety of indicators that can best capture the broader dynamic aspects that interventions are trying to address. Moreover, the evaluators usually need an extensive knowledge of the society in which the intervention is taking place in order to select indicators that best measure the effectiveness of peacebuilding programs. Good indicators should be decided upon by different stakeholders with varied interests. It is fine to have competing indicators since there is no single socio-political reality or impact. While indicators are often context-specific without a standardized list of peacebuilding indicators, UNDP Myanmar identified two key categories of indicators for PCIA:

- **Conflict/context indicators** monitor the main issues and drivers of conflict identified in the conflict analysis. This helps to keep conflict analysis up to date and monitor conflict sensitivity issues linked to the project. It can measure change in a broader context or at a more project-specific level.
- **Interaction indicators** allow for a more dynamic and continuous form of monitoring to provide up-to-date information from which to measure the interaction between the context and the intervention. After identifying interaction points where planned activities might interact or connect with issues and challenges highlighted in conflict analysis, create indicators that can measure the effect of intervention on conflict context and vice versa.
  - For example: number of violent disputes between groups in target villages before and after project, perceptions of trust before and after project, days that staff are unable to travel due to security concerns.

*To find the relevant context and interaction indicators for the intervention that you are monitoring, look to [“Indicators for local and community development programming in UNDP Myanmar”](#) p. 16-21 for some examples.*

Instead of conducting a one-shot assessment of indicators, it is best to monitor the context and interaction of indicators periodically. This way, when there is an indication of the activities doing harm or if there is an opportunity to support peace, project managers can adjust the planned activities accordingly. After collecting information on the context and interaction indicators, note the difference between early and later stage assessments to identify the wider impacts of intervention on the context and vice versa.

To overcome the one-dimensional nature of using variable-based approaches, it is recommended that you employ mixed methods. This involves triangulating data, methods, and researchers from different disciplines whenever possible. Triangulation may help create better knowledge on complex situations. However, when presented conflicting information from different parties, this may be an opportunity to investigate further.

### **Process based approaches to PCIAs**

PCIAs ask broad-level questions such as:

- i. If projects produced substantial or politically significant changes in access to material and non-material resources.
- ii. how socio-economic tensions are created, exacerbated or mitigated.
- iii. The kinds of challenges or changes in content of control over existing economic or social systems.

For these complex questions it is hard to attribute an impact to a single cause produced by a project. Moreover, a project could be doing exactly what it set out to do but unexpected or unplanned impacts could prevent it from achieving its peacebuilding objectives. As such, social science methodologies in these circumstances may not be able to meet the demands for impact assessments with direct causal inferences. Therefore, process-based approaches may be more useful for evaluating the broader and unintended and unanticipated impacts that come with peacebuilding projects.

### **Participatory Approach for PCIAs**

PCIAs need to be a dialogue between the assessment team, local stakeholders and implementing agencies. Therefore, employing participatory approaches help to empower locally-based evaluators and give those affected by conflict a central role in planning, monitoring, and evaluating projects for conflict sensitivity. Since there will often be multiple and conflicting perspectives to consider, ensure a level of diversity in participants chosen to contribute. If divisions exist, consider conducting separate consultations and using language that avoids conflict (i.e. using context analysis instead of conflict analysis). This approach is useful in the same ways as process-based approaches and gives stakeholder more chance to give their inputs and subsequently trust the evaluation. However, it is more suitable at a micro rather than a macro

level. Another disadvantage is that giving participants a larger role in designing the evaluation could serve as a conflict of interest in polarised situations.

Participatory approaches are also important because they can serve as an opportunity for institutional learning on working in conflict situations, which are of greater interest to many agencies. A common framework should be developed locally and applied by all donors operating in the country rather than imposed by donors.

## Managing Evaluations

### Contracting Evaluation

Calling a contractor to conduct an evaluation can bring in external experts with in-depth knowledge of an area but has the disadvantage of not increasing in-house evaluation capacity. Contractors might also lack local knowledge - so if used, they must be given the opportunity (time, allowance, etc.) to gain necessary familiarity. The successful candidate is selected by a panel of people with evaluation knowledge and experience.

### Roles and responsibilities of different players

The **Main Client** requests and usually funds the evaluation. They should discuss with you the nature and context of request, needs, timing, communication schedule and frequency. They lead and ensure the development of a timely implementation of an evaluation plan. The client appoints an evaluation manager to oversee the day-to-day activities.

An **Evaluation Team** consists of a manager and different specialists. Tasks can be divided using task maps to shape individual assignments. **Composition** should be a balance of different genders, ethnicities, ages, and other identity markers.

The **Manager** manages all stages of the evaluation process from evaluability assessment to dissemination results and follow up. Along with recruiting, managing, and overseeing the evaluation team, the manager also establishes first contact with key stakeholders throughout the process and connects relevant evaluators to them in order to ensure a fully inclusive and transparent approach to the evaluation. They review and comment on draft evaluation reports, circulate draft and final evaluation reports. Finally, managers also facilitate knowledge-sharing and use of findings in programming and decision-making.

**Evaluators** come from different areas of expertise, be they survey designers, sampling experts, or gender advisors. As an evaluator, you will report to management on compliance with the evaluation plan, completion of management responses and key actions and quality assessment results. You will contribute to and review the draft and final evaluation reports.

In a participatory evaluation, stakeholders should be consulted at different stages of the evaluation process. Stakeholders who are consulted for the evaluation should involve critics of the evaluation or intervention as well. To reach out to these stakeholders, you can build trust by

sharing information with them from the onset, working through intermediaries, starting early and broadly.

#### Writing a Term of Reference to guide an evaluation

Terms of references (TOR) should be explicit and focused, with a clear mandate about what to evaluate and why. Make sure to take into account real-world constraints such as budget, skill capacities and timescales. Within these constraints, the evaluation team should retain enough flexibility to determine best methods and tools. See [Appendix F](#) for a sample of a Term of Reference template.

## Ethical Considerations

Ethical considerations in research refer to professional codes that guide the conduct of research. This section will explain how to create an accountability framework that guides all steps of the research, ethical considerations for research design, and ways to recognize research corruption.

### Creating an Accountability Framework

Accountability is the characteristic of a relationship based on shared research norms or expectations. Those who hold researchers accountable include not only the donor(s) but also the vast network of stakeholders invested in the research results which include the community being researched and the policy community. Being accountable to the researched community aligns with the principle of ‘do no harm’ during and after the research. This principle suggests that the research should serve and promote the welfare of a target population and make sure that it does not have harmful consequences on respondents. Meanwhile the policy-community could have political interests in obtaining specific answers from the research. It is important for you to understand and balance the overlap in expectations, needs, and requirements among different stakeholders. This is why mapping out relevant stakeholders in advance using a stakeholder analysis is so important.

Upon taking into consideration multiple stakeholders’ expectations toward the research, you can form an accountability framework using contracts or an ethics review processes. Whatever the format, these frameworks should include measures of **answerability** and **enforceability**. Being answerable means providing enough information about the outcome of research to participants so that they can evaluate the risks and benefits of the survey for themselves and their communities (Berdahl & Archer 2015; Halperin & Heath 2017). This could take the form of an ethics review process where stakeholders ask you questions to judge your behaviour against shared expectations. Ethical reviews often consist of a series of questions and considerations that you reflect on and gain approval for prior to conducting research. While ethical reviews can be onerous and time consuming, they can also be helpful in uncovering a range of issues you might not have considered otherwise. Whichever format is chosen, you should provide information and respond to questions in a transparent, timely, and accurate manner. **Enforceability** involves imposing consequences such as sanctions if you have not acted according to shared expectations.

You need to clearly indicate which stakeholders from your stakeholder analysis you are answerable to and who has the power of enforcing accountability.

## Ethical Considerations for Research Design

Research organisations should consider ways in which ethical and safeguarding reviews are built into their research at design stages. Doing so should ensure that safeguarding of research participants and yourself as a researcher is properly considered.

These considerations should help you form strong working relationships with the communities that you are working with:

- **Clarify your role to communities and relevant actors through appropriate methods.** Your presence could raise expectations in the community. You might contribute to existing conflict depending on how you are viewed by targeted communities. For example, the organization you come from might not be well-trusted or well-known within the community. Therefore, it is important to consider who is conducting assessments and how they are introduced to the researched community.
- **Clarify justification for selecting the researched community to relevant stakeholders.** Deciding who gets consulted among groups close to chosen group(s) is a sensitive matter that may risk increasing pre-existing divisions. Make sure that the selection criteria and partner agreements with codes of conduct for consultations are transparent and shared among targeted and non-targeted groups. When consulting non-targeted groups that are similar to the targeted groups it might also help to explain the research's purpose. This will reduce perceptions of bias and partiality amongst groups that might be experiencing tensions.

These considerations are necessary for data collection

- **Ensure anonymity and confidentiality of research participants**

In this part, respondents are assured that their personal information will be kept separate from the data and the data will be managed in a way that their identity will be kept confidential and will not be released as part of the data. Confidentiality in focus groups research is very challenging as both external and internal confidentiality must be considered and respected. External confidentiality, which relates to the disclosure of participants' information as part of the data, can be respected by providing a statement where you guarantee that you will preserve participants' anonymity. However, internal confidentiality entirely relates to members or participants of focus groups and the researcher has no control over the disclosure of information by those individuals. Therefore, to respect internal confidentiality, it is recommended that "focus groups should be composed of individuals previously unknown to each other" (Sim & Waterfield 2019). When participants are selected from the same community, whether previously known or not, they have to make an agreement that no personal information said in the group is to go outside the group - otherwise participants will not fully engage in group discussions.

- **Obtain research participants' informed consent**

The most important code of conduct, the “informed consent” rule recommends that a participant must be given full and honest information about the research to be undertaken, its objectives, and potential risks and benefits. You will need to provide potential participants with information about aspects, purposes, and consequences of the research. The provision of full research information to respondents is easy in quantitative and more structured studies that include pre-determined questions and a specific scope that does not allow participants to go beyond the questionnaire. But in a focus group, where the research design and questions are not pre-specified and most information is expected to emerge from the research process, it is difficult to provide all the information that a participant needs before agreeing to participate in a focus group. Therefore, it is important to provide the respondents with as much information as possible about aspects and purposes of the research and also about the nature of focus groups and the possibility of the emergence of unexpected information and structures from the process of research. Provision of such information helps participants to have an image of not only the research topic and aims but also the research environment and structure. Along with providing information, you need to inform participants that although their continued participation is assumed, they can withdraw their consent at any time (Sim & Waterfield 2019).

Based on this information and your guarantees, the respondents should be allowed to voluntarily decide to either give consent or refuse to participate. This principle will be violated when participants' decision for consent is influenced by your authority as a researcher or any pre-existing relationship you share with the participant (i.e. as a teacher and student, physician and patient, or employer and employee). Participants can officially declare consent by signing an informed consent form along with the researcher. A typical consent form includes a number of ethical and structural components concerning the research aims and the nature of participation that are highlighted in the example provided by the University of Edinburgh ([see Appendix E](#)).

- **Create parameters for vulnerable populations to participate**

Stakeholders often come from vulnerable, traumatized, disenfranchised or otherwise disempowered populations. Responsible organizations have to make sure staff do no harm to vulnerable adults and children and do not expose them to risk of harm and abuse beyond ‘compliance’ to ethical standards. **Vulnerability** is the consequence of deep-rooted inequities that divide societies along the basis of gender, age, class, and other social groups.

When creating question messaging to vulnerable populations, you should be aware of broader socio-political problems and how asking questions may challenge an unjust status quo or groups in power. Therefore, anonymity and confidentiality are especially important for gathering data among vulnerable populations who might fear reprisal for participating. This can be done by choosing data collection methods that ensure the highest level of anonymity such as individual interviews and anonymous surveys over

others such as focus groups. Another reason why individual interviews are preferred over focus groups is to avoid power relations that could misconstrue data in large group discussions. If you have to conduct focus group discussions, consider dividing groups by identity indicators such as gender, income level, age, marital status, ethnic group. Another safeguard involves distinguishing between internal information that should only be shared among research staff for learning purposes and external information that can be shared more broadly with donors.

- **Disaggregate data by vulnerability and conflict-sensitive groups**  
For research to better reflect the situation of marginalized groups, disaggregate data by identity indicators such as gender, income level, age, marital status, ethnic group. Similarly, data in environments affected by conflict should be disaggregated by communities that are known to have tensions between them.
- **Conflict-sensitive data collection:**  
When data is difficult to collect in conflict-affected areas, explore alternative data sources using technology or social media tools for remote data gathering

## Recognizing Research Corruption

Different stakeholders may have different interests in influencing research. Research with political aspects can be especially corruptible. At worst, evaluators twist the truth to produce positive findings as a result of conflict of interest or other perceived payoffs or penalties. At best, prejudices and biases can shade evaluation findings. You need to be aware if and when stakeholders are trying to influence results without regard for the research's validity. This might not only manifest in explicit forms such as bribes but also more subtle ways that are not immediately recognizable. For example, you might unconsciously give more powerful voices higher priority.

Here are some ways to recognize influences. At the beginning, stakeholders might declare certain questions or areas of evaluation off limits without substantial information. During the evaluation, they may delay the process, omit information, or come up with new data towards the end of the evaluation. Finally, at the end, stakeholders could modify results before a report's release, suppress certain findings, or deny or downplay identified problems. You should not investigate allegations of corruption but rather report to the appropriate authority for investigation.

For more information, refer to the ethical considerations and responsibilities in social research outlined and described by the ICC/ESOMAR and the World Association for Public Opinion Research (see ICC/ESOMAR 2016; WAPOR 2011).

## Part 2: Methodologies for Research and Evaluation

### Sampling

#### Overview

Sampling refers to methods of selecting a representative sample from a target population. A sample is a microcosm of a population. This section describes various methods of sampling, the scope, and purpose of each method, limitations, ethical considerations, examples, and the application/relevance of each method to a particular research design. For qualitative research, the most popular sampling techniques are described. For quantitative research, four sampling methods are described and their logic and purpose of application are discussed.

#### Keywords

Target Population, Sample, Sampling Frame, Sampling Unit, Sample Size, non-Probability Sampling; Probability Sampling; Simple Random Sampling; Stratified Sampling; Systematic Sampling; Cluster Sampling; Sample Size.

#### Concept Definitions and Terms

In social research, the ideal method is to study an entire target population, which is rarely possible in practice. Therefore, researchers use different methods to take a sample of the target population for the study (Acharya et al., 2013; Daniel 2012). The method of selecting such samples is known as “sampling.” Sampling allows a researcher to study a representative number of respondents to know about their preferences, knowledge, opinion or attitudes concerning a particular policy or a social problem. In general, taking a sample of a population requires the recognition of four criteria including a *target population*, *unit of analysis*, *characteristics of a sample*, and *a sampling frame*.

**Target Population and Unit of Analysis:** Target population refers to the total population of interest for the study. A target population could include different units of analysis including individual Afghan citizens or Afghans, university students, newspaper articles, etc. The identification of the target population and units of analysis depend on research questions and the purpose of research. For example, if researchers want to know about Afghans' responses, the target population is Afghan citizens and each citizen is a unit of analysis, but if the purpose of the research is to study the knowledge or attitudes of university students, the target population will be all university students, with individual students being the units of analysis. Identification of the unit of analysis is important in sampling from a larger population (Berdahl & Archer 2015).

**Sample** refers to a subset of the *units of analysis* that a researcher selects from a target population. When the unit of analysis is individual citizens and the target population is citizens of a country, the sample is the subset of individual citizens (Berdahl & Archer 2015).

**A sample frame** is based on a list of all units of analysis in a population of interest. This usually includes a list of units of analysis and their addresses/contacts from where a representative sample can be drawn. For example, if your unit of analysis is adult Afghan citizens, you need to provide a list of adult citizens in each of the provinces or districts that are accessible for interview or survey. If your target population is members of Afghan political parties, but you cannot access members of all parties, a practical sampling frame could include a certain number of a few political parties that are willing to participate in your research.

For example, in a survey of people's perspectives on political settlement and post-conflict order, the researcher provided a sampling frame of 18.6 million adult population in 34 provinces of Afghanistan from which 1500 individuals in 34 provincial centers were selected as the sample (Ibrahimi 2020). Relying on the same data in 2018, the IWA identified a sampling frame of 13.6 million population and selected a sample of 8130 individuals in all provinces of the country (McDevitt & Adib 2018). Even if the unit of analysis is different from individual citizens, the sampling frame must follow the same logic. For example, if a researcher considers a sampling frame of state university students in Afghanistan in 2021, the framing list must include the total population of all registered students in state universities. The research sample will be selected from the list of registered students in the sampling frame by using a particular sampling method.

### Sampling Methods

In social sciences, we regularly use two methods of sampling including probability and non-probability methods. Non-probability methods are commonly used in qualitative research such as individual interviews, focus groups, ethnography, etc. The most common non-probability sampling methods include purposive sampling, convenience sampling, snowball sampling, and quota sampling.

In purposive sampling, which is also known as selective sampling, respondents are selected with the assumption that they can provide detailed information about the research issue. The criteria of selected respondents are determined by the researcher. In the convenience sampling method, the sample is selected on the basis of accessibility or availability to researcher. This method is applied in studying populations in areas that are accessible and familiar to the researcher. Snowball sampling is another method of selecting a sample from a population where the sample group grows like a rolling snowball. This is a sampling technique that a researcher uses in difficult environments such as conflict zones or critical areas. The sampling starts with researcher's first contacts with a few individuals who are relevant to the topic of interest and uses them to establish contact with other possible respondents. In this process, one individual refers the researcher to another, who in turn provides the address of a third respondent, and so on, until the required sample size is identified and interviewed. Finally, quota sampling is another method of non-probability sampling that is useful when a researcher is interested in studying specific strata such as age categories, income categories, etc., within a population.

Non-probability sampling methods are not applicable in quantitative research for which we usually use *probability sampling* techniques. In probability sampling, the sample or the number of assigned respondents to a research is selected by using a *random* process. The random process

ensures that each unit in the population has an equal chance of being selected in the sample. This is to help ensure that the results obtained from studying the sample reflects the characteristics of the population as a whole. In general, there are four methods of probability sampling that follow the random selection criteria: simple random sampling, systematic sampling, stratified random sampling, and cluster sampling.

**Simple random sampling** is a basic sampling technique where each unit of a target population has an equal chance of being included in the sample. Researchers select each unit of the sample or each individual entirely by chance and randomly. Simple random sampling helps ensure that the sample is representative of the demographic combination of a target population and it is used to make statistical inferences about a population. (Easton & McColl 1997).

Practical steps of applying simple random sampling includes: 1), define the population by providing a complete list of the target population, i.e. citizens of a country, inhabitants of a province, or participants of a conference, etc.; and 2), calculate the size of your sample. The common method for selecting a representative sample is using your desired confidence interval and margin of error. The margin of error is based on the assumption that when the error in research is inevitable, how much error does a researcher allow in a study. Researchers, regularly select samples with 5% margin of error and a 95% confidence interval. This means a researcher sets a framework for sample selection that explains a 5% difference/error between the result that he/she obtains from the sample and the hypothetical result from the general population. The 95% confidence interval tells you that the actual or the population mean falls into your 5% margin of error and you can say it with 95% confidence. Likewise, if you decide a different margin of error, for instance, a 3% margin of error with a 95% confidence interval, you are expecting that the difference between the result that you obtain from the sample and the result that you could obtain from the entire target population is less than 3% and you can say it with 95% confidence, and so on (Acharya et al., 2013; Gelman 2004; Thomas 2020).

Besides this, the most common margins of error are 5% and 3% and the most common confidence intervals are 90%, 95%, and 99%. Any margin of error and confidence interval that a researcher selects determines their sample size. For a higher confidence interval and a smaller margin of error, you need a larger sample size and vice versa (Daniel 2012; Berdahl & Archer 2015; Determining). Boyd (2006) provides a comprehensive table for how to determine the sample size for a target population with various margins of error and confidence intervals. You can also determine your sample size for various target populations by using the Sample Size Calculator (2020) the details of which are listed in the references.

An example of simple random sampling would be the study of 200 participants in a conference when a researcher wants to study a sample of 50 participants. Here, the 200 participants are defined as the population and the 50 individuals are determined as the number being required for the sample. With having the complete list of the target population (200 individuals) and a determined sample (50 individuals), the researcher can randomly select the sample by using the *random number* method or *lottery* method (Acharya et al., 2013 Thomas 2020).

Simple random sampling is useful when stratification or systematic sampling of a population is difficult or not possible. Simple random sampling is also useful for telephone, email, or other methods of data collection that requires using electronic devices. Moreover, simple sampling is useful for creating and studying larger samples. Another advantage of simple sampling is that it allows for selecting a study sample without considering respondents' level of knowledge and other complex demographic features. Therefore, data analysis for generalization from such a sample is not difficult.

**Stratified sampling** is a sample selection method from a target population that is divided into clear subpopulations or strata that share common characteristics. Therefore, a stratified sample is obtained by taking samples from each stratum or sub-group of a target population such as province, district, ethnicity, political party, income category, etc. When you sample a population with several strata and your sample is small, stratification helps increase the likelihood of representativeness. In this case, stratified sampling helps 1), use a smaller sample for generalization and 2), resolve the problem of inadequate sampling of some subgroups within the sample (i.e. while some subgroups are oversampled, the final statistics is expected to be proportionally adjusted in order to obtain whole population estimates).

It is also important to note that stratified sampling techniques are usually used when the population is heterogeneous, or dissimilar (Easton & McColl 1997). This method provides an estimation of each subgroup and assures the representation of all groups in the target population. However, the provision of a sampling frame in this method requires accurate information about the proportions of each subgroup or stratum (Acharya et al., 2013). Therefore, without the researcher's in-depth and detailed knowledge about characteristics, combinations, and proportions of subgroups in a target population, it is not easy to provide an accurate and representative sample through stratified sampling.

In general, stratified sampling is usually used for selecting smaller samples. An example of this method can be the selection of 30 Wolesi Jirga members (Afghan MPs) with the assumption that they represent Afghanistan's nine major regions. In this example, the list of all MPs of the nine provinces provides the sample frame. Then, you need to divide the MPs of each province by the total number of MPs (114) in the sample frame. The result will be the proportion of each province in the Wolesi Jirga. Finally, for obtaining a stratified sample, you need to multiply the expected sample (i.e., 30 members) with the proportion of each province's MPs. The result will look like the following table.

*Table 7: Stratified Sample of 30 MPs of Afghanistan's Nine Major Provinces*

Province	Population of MPs	Proportion of MPs	Stratified Sample of 30 MPs
Kabul	33	0.289	$30(0.289) = 9$
Herat	17	0.149	$30(0.149) = 4$
Balkh	11	0.096	$30(0.096) = 3$

Kandahar	11	0.096	$30(0.096) = 3$
Nangarhar	14	0.122	$30(0.122) = 4$
Bamyan	4	0.035	$30(0.035) = 1$
Badakhshan	9	0.078	$30(0.078) = 2$
Parwan	6	0.053	$30(0.053) = 2$
Paktia	9	0.078	$30(0.078) = 2$
	<b>114</b>	<b>1</b>	<b>30</b>

Data Source: Wolesi Jirga Website.

**Systematic sampling** is a method of selecting a random sample from a list of all possible units in a target population. In this method, a researcher needs to first identify a sample frame or a list of possible units of analysis, and, next, he/she should start sampling by selecting the first unit from the sample frame randomly. The first unit will be selected by a simple logic: that is, divide the number of units in your sample frame by the number of the desired sample. The result will provide your sample selection criterion as presented in the figure 1:

Figure 6: Selection Criteria Equation

$$\frac{\text{Sampling Frame}}{\text{Desired Sample}} = \text{Selection Criterion}$$

After clarifying your sampling criterion, you should continue selecting your subsequent units by using a systematic/periodical method of random selection. For example, if one million households in Kabul province include your sample frame and you aim to systematically select a sample of five thousand households for your study, the systematic sampling of the five thousand households will be done as follows: divide the sample frame (1,000,000 households) by the expected sample (5,000 households):

$1,000,000/5,000=200$ . The number 200 is your selection criterion of the sampling interval. In other words, 200 is the first number that you can start with your sample selection of target households. Start with a case between 1 to 200 and choose every 200<sup>th</sup> cases. For example, select the 200<sup>th</sup> case as the first case of the sample, the second case would be the 200<sup>th</sup> case after 200 which will be the 400<sup>th</sup> case, then the 600<sup>th</sup> case, and so on.

Systematic sampling is useful for both small and large samples depending on research design and purpose. Systematic sampling is simple to draw and very easy to verify. One disadvantage of this method is that it can introduce bias where the systematic approach picks up some systematic variation in the listed population. In this method, the first unit is selected by a different method

than the rest of the units. The first unit is selected by using a probability method while the subsequent units are selected systematically when all units between two systematic cases are given a zero chance to be selected in the sample (Acharya et al., 2013; Daniel 2012).

**Cluster sampling** is a two-step process of selecting a sample from a target population. First, the researcher divides the population into clusters or groups, usually geographic areas such as provinces, districts, or villages, and next, they select a sample from each cluster by using simple or systematic sampling methods. This is a useful method of sampling when the target population is widely scattered like the population of a country and divided by specific units such as provinces, districts, or other geographic clusters and the method of data collection involves face-to-face contact with each member of the sample.

Therefore, cluster sampling is most practical to be used in national surveys that require a large sample size (Acharya et al., 2013). In the past two decades, several national surveys are conducted by drawing on cluster sampling in Afghanistan. For an example of the application of this method in Afghanistan see the Integrity Watch's (2018) survey which has interviewed over 8000 respondents based on smaller clusters, districts. This survey first identifies 140 districts in 34 provinces of Afghanistan, and next it provided a sampling frame of villages in rural areas and *nahias* or divisions in urban areas from which respondents were selected by using a simple random sampling method (see McDevitt & Adib 2018).

### **Sample Size**

Regardless of what method of sampling you use, it is important to know how many people should be included as respondents in research and why. However, all sampling methods in identifying a sample size, follow two criteria: the sample size depends on the diversity of the population and the number of variables being studied. Therefore, using a higher number of variables for studying a population with greater diversity requires a larger sample. As a result, the question is: how to select a correct sample size? The answer is explained in the following.

After selecting research variables and specifying the sample frame or the list of a population that fits the research scope, sample selection can be determined by taking into consideration two factors: *margin of error* and *confidence interval*. Details about selecting the required margin of error and confidence interval are explained in "simple random sampling" section.

### **Conceptual Framework**

In quantitative research, sample selection depends on the conceptual framework of the research. As a result, a sampling frame requires the identification of independent, dependent, and intervening variables, and the specification of the nature of the relationship between them. The sample selection must cover the number of respondents that know the variables of the study and are eligible to respond to the kind of relationship among them.

For example, in a study of post-conflict order in Afghanistan where Loya jirga, local councils, religious mechanisms, elections, and ethnic politics are included as study variables, a researcher needs to select the sample with the assumption that the respondents know about those variables

(see Ibrahim 2020). Otherwise, the sample frame will not fit the conceptual framework which makes obtaining an accurate result difficult.

One of the best practices for fitting the sample to the conceptual framework is to increase the sample size when the number of variables is increased and vice versa. In other words, the inclusion of a higher number of variables in research requires a larger sample size, because it is expected that parts of the sample might not respond to certain questions or state “don’t know” to certain questions. Therefore, to decrease selection bias and useless data in research with so many independent and intervening variables, the best practice is to enlarge the sample size.

### **Interpreting results**

Interpreting results in quantitative research requires statistical methods of data analysis. The most common methods of data processing and analysis include STATA and SPSS. These two software products help produce descriptive data and inferential data and facilitate bivariate analysis and multivariate analyses (Haan & Godley 2017).

The descriptive data provides numbers and tables that describe the characteristics of the sample and its subgroups. For example, how many men vs. women would like to participate in the forthcoming election or the percentage of the rural population’s interest in agricultural growth compared to those of the urban populations.

By contrast, inferential data provides estimates of characteristics of the population, based on data from the sample. For example, the average age of respondents’ interest in democracy or the average support of women for freedom of expression. A bivariate analysis draws comparisons between groups, or examines the relationship between two variables. For example, how people’s level of religiosity affects their level of support for democracy or how their province of residence influences their level of support for the central government.

Finally, a multivariate analysis such as multiple regression describes the level of association between several independent variables and a dependent variable, as a group. It will also allow researchers to assess the particular contributions of each independent variable to the outcome. For example, when you are testing the relationship between respondents’ province of residence and their level of support for central government, multivariate analysis helps you to test the impact of other factors such as the respondents’ level of education, their income, ethnicity, and other factors on the relationship between independent and dependent variables. Factor analysis is another commonly used multivariate procedure which allows for reducing a large number of variables into a few interpretable factors (Young & Pearce 2013).

## **Surveys**

### **Overview**

Survey is a method of data collection from a sample by using a questionnaire and structured interviews. Questionnaire refers to a list of mostly closed-ended questions that include

categorized responses. Survey results provide general conclusions and standards about social knowledge and attitudes. Such conclusions become sources of analysis and judgment for policy-makers and the general public about specific issues, policies, behaviors, and decisions. This section describes survey, the process of applying this particular method, survey questions and questionnaires, data analysis strategies, and the policy implication of survey methods.

#### Keywords

Survey, survey design, questionnaire, questionnaire design, best practice for policy, ethical considerations.

#### Concept Definitions and Terms

Survey refers to both a research method and a method of data collection. When you use survey as a research method it includes all processes of inquiry from research design to data collection and data analysis. This section uses survey in its broader term as a research method and treats data collection as an element of this method.

Survey in its broader term has three characteristics. First, it is a quantitative (and qualitative) method of inquiry for studying different aspects of a target population through examining the relationship between a set of variables; second, it is a method of data collection that uses structured questionnaires for collecting data from a sample; third, it requires a representative sample from which a researcher aims to generalize about the target population. Survey is a useful method of inquiry for studying a quantifiable sample, its demographic composition, its opinion, preferences, behavior, etc. (Check & Schutt 2012; Glasow 2005).

Application of survey research requires a scope of the inquiry or the identification of the study variables, the specification of a sample that a researcher wants to know about their responses about the relationship between variables that are identified in the earlier stage, provision of a list of questions in a questionnaire that help operationalize the inquiry, a data collection method, and a data analysis method.

Depending on the purpose, scope, and design of research, researchers can design their survey questions and the questionnaire in a variety of ways. However, the most commonly used questions in survey research are closed-ended. There are two reasons for relying on closed-ended questions. First, they facilitate and accelerate the process of data collection by providing respondents with limited categories of responses for each question. Second, this method eases data analysis by providing comparable answers. By contrast, when open-ended questions are used in survey research, data analysis is more time consuming and their combination with quantitative data can be quite difficult (Berdahl & Archer 2015; Check & Schutt 2012; Halperin & Heath 2017; Singleton & Straits 2009).

Overall, survey research starts with a conceptual framework that requires the scope of research and the study variables, a questionnaire design, a clear sampling agenda, fieldwork, and data

analysis. The remainder of this section explains these elements along with their policy implications and practices.

### **1. Conceptual Framework**

The first step for conducting survey research is the identification of the dependent and independent variables that a researcher wants to study. For example, if the survey is supposed to study causes of war from a target population's perspective, the researcher has to list several variables as causes of war when the war itself can be treated as the dependent variable. Description of the dependent and independent variables in the earlier stage of an inquiry helps clarify the scope of research and facilitates the design of questionnaire. Thus, before starting the practical steps of a survey, a researcher needs a conceptual framework that describes the study variables and their hypothetical relationship.

Without clarifying the dependent variable(s) and identifying relevant independent variables about which you want to know the idea or preferences of respondents, it is not possible to develop a rational questionnaire. The choice of dependent and independent variables requires in-depth pre-test inquiry including literature review or empirical observation to clarify what has been done about the problem of interest and what is worth researching. A literature review is also helpful in identifying relevant independent variables as instruments of inquiry.

There are many examples, that provide a clear image of how to provide a conceptual framework and develop a questionnaire that includes a list of questions related to the study variables for this topic. In recent research on conflict resolution in Afghanistan, Ibrahimi (2020), identified three dependent variables including "political settlement," "post-conflict order," and political legitimacy in Afghanistan and examined people's views and preferences about the three topics. These dependent variables, initially, emerged from the following research questions:

1. Which mechanism of political settlement do people prefer?
2. What is the people's preferred political system in a post-conflict setting?
3. How do people evaluate legitimacy of government in Afghanistan?

Following the construction of research questions and identification of the dependent variables, four independent variables were identified for studying the first dependent variable (i.e., political settlement) including elections, power-sharing government, decentralization of power, and interim government. This list of independent variables led to a series of related questions in the questionnaire. The same method was applied for addressing the second and third questions on political order and mechanisms of political legitimation in the country. Two independent variables including democracy and autocracy were identified for political order and legitimacy was measured by two variables including traditional and rational-legal methods of legitimation. Similar to the first variable, relevant questions to these independent variables were added in the questionnaire. As a result, a questionnaire with a list of closed-ended questions around the identified variables was created. Details and techniques of creating and ordering the questionnaire will be discussed in part 3 of this section.

In general, developing a conceptual framework helps clarify the scope of research and provision of the questionnaire and its instruments. Without having the dependent and independent variables clarified, it is almost impossible to develop an accurate and clear research agenda, research scope and purpose, and a rational questionnaire. Therefore, the best strategy for applying survey research is to start with identifying your variables and an in-depth inquiry or observation to identify their elements to create a list of researchable questions. Every single question in the questionnaire has to be logically connected to the study variables and represent an aspect of them.

## **2. Survey Sampling & Survey Design**

Depending on the topic of interest, the target population, the unit of analysis, the scope of research, and the purpose of inquiry, any of the four sampling methods that are described in the Sampling section can be used as a method of survey sampling.

Besides sampling, a survey requires a design for research application or the choice of a medium through which the survey will be administered. Both written and oral methods can be distributed to respondents by two general methods: mail and email (or web survey) depending on respondents' access to any of these items. Besides the written survey, the survey can also be conducted orally by using the conventional face-to-face survey methods when the researcher hands out a questionnaire to participants and wait while they fill it out. Field refers to the area or geography of research including cities, villages, communities, or any other unit of a society that is identified by the researcher as the areas of study. Verbal surveys can also be conducted by telephone where the researcher reads every question of the questionnaire and the answer choices to the single respondents and record their responses (Glasow 2005).

The face-to-face survey provides the most reliable data and is “well suited to long or complex questionnaires and for reaching the correct respondents” (Glasow 2005). But this design is expensive and difficult to apply in unsuitable conditions such as conflict zones, war-affected areas, or areas under the control of military or political groups that are against the inquiry of certain topics or data collection from a particular population or area. In this situation, researchers aim to use alternative measures such as snowball sampling or purposeful sampling which mostly lead to qualitative interviews rather than a quantitative survey (Cronin-Furman & Lake 2018; Cohen & Arieli 2011).

Moreover, mail surveys are not a practical method of data collection in countries such as Afghanistan, where households are not systematically addressed. By contrast, telephone and email surveys are easy and inexpensive to conduct in such countries. But the reliability of the results of telephone and email surveys can be affected by coverage biases where parts of the population, particularly in rural areas, do not have access to a telephone number or an email address. Therefore, in applying any survey method, a reliable inquiry must take into consideration context related constraints and the living conditions of the target population.

### **3. Questionnaire: what to ask; how to ask?**

The provision of a questionnaire requires a list of closed-ended questions and a rationale about their wording and order. This section explains survey question, question wording and format, and question order as key elements of designing a questionnaire.

Survey questions are regularly closed-ended with answer choices. Closed-ended questions are easy to answer – particularly when the choices offered fit the respondent's views or experience –, easy to code, easy for data entry, and easy to analyze. But they regularly simplify complex issues and are influenced by the researcher's knowledge of the topic and his/her determination of a limited number of choices for respondents. In other words, such questions leave respondents with a limited list of responses that are provided by the researcher. Therefore, when closed-ended questions are limited to the analytical framework and knowledge of a researcher, the provision of questionnaire must be done with caution and by a research team that is highly familiar with both the scope of possible responses in a target population and the cultural or political context. Moreover, to develop a rigorous questionnaire, it is important to send the first and later drafts for peer-review and fact check by field experts and to a sample of potential respondents, to understand how they might interpret and respond to the questions and the answer choices offered.

A survey questionnaire is regularly divided into two general parts. First, the part belonging to demographic questions that examine respondents' demographic characteristics such as age, sex, education, income, political affiliation, ethnicity, and the like. Second, the part belonging to informative or neutral questions, and key or main questions. While informative or neutral questions are used for starting the interview with respondents or for obtaining their level of information about issues surrounding the survey topic, the key questions are exactly about the topic of research and fall into the conceptual framework of the research. Key questions, in general, examine respondents' knowledge, attitudes, behaviors, and preferences concerning the survey topics.

### **4. Question Wording, Order, and Format**

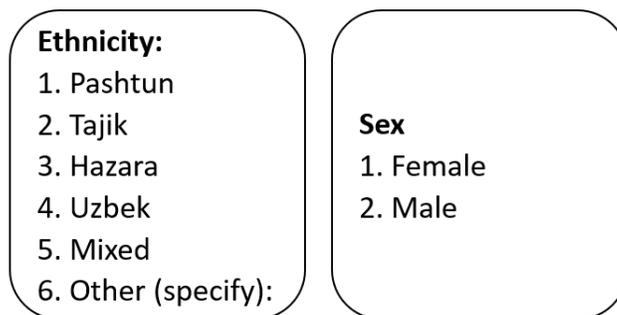
Question wording, question order, and question format are elements that must be considered for designing a questionnaire. Question wording should follow two general criteria including single quality criterion and clarity of questions. Concerning the single quality criterion, each question must measure only one quality. For example, when you are planning to survey customers' perspectives about a restaurant's price and services, create one question about the price and one question about the services. The inclusion of both qualities in a single question confuses and produces unreliable data. Concerning the clarity of questions, a question should not be ambiguous or open to multiple interpretations. This means every question should be measured by clear indicators that are easily understandable to a target population. Third, every question should be commonly understood by all respondents. Therefore, it is important to make sure that all respondents, regardless of their demographic differences, have a common understanding of the question and its measures or indicators (Berdahl & Archer 2015). Question wording influences respondents' choices and the reliability of data.

Regarding question order, taking the three types of survey questions (i.e., demographic, informative, and key questions) into consideration, questionnaires regularly start with informative or neutral questions. For example, in a survey on the peace process in Afghanistan, an informative question can be used as follows: Do you know about Afghanistan’s peace process (Yes/No). Starting the questionnaire with such questions creates a comfortable environment for a respondent preparing him/her for answering the key questions. After one or a few neutral questions, the questionnaire proceeds with key questions that form the basis of the survey. Following the key questions, the questionnaire usually ends with demographic questions. The reason for putting demographic questions stems from the assumption that asking questions about details of a respondent’s personal life can make him/her uncomfortable and it could influence their choices while answering the key questions. Overall, the general rule for ordering the questions of a questionnaire is to start with neutral/informative questions, proceed with the key questions, and end with demographic questions (Berdahl & Archer 2015; Halperin & Heath 2017).

Finally, question format is another important requirement that must be considered for designing a survey questionnaire with multiple-choice questions. Multiple choice questions regularly take three formats: nominal, ordinal, and interval/ratio questions.

**Nominal questions** refer to categorical questions about names, places, categories, and demographics. Every response or category carries a single value that cannot be ranked from highest to lowest or from best to worst. Every question requires all possible responses to be listed, and one or two responses beyond the recognized list such as “other” or “I don’t know,” etc. Nominal questions produce the kind of data that can be analyzed with frequency distribution and univariate statistics. These questions can also be used as independent variables in bivariate or multivariate analyses. Two examples of nominal questions are as follows (Ibrahimi 2020):

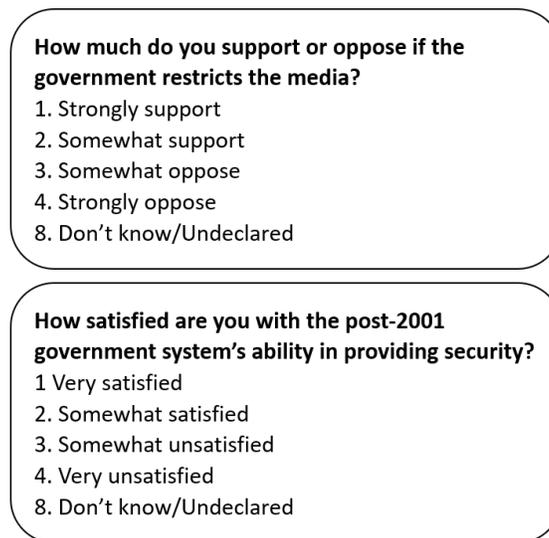
*Figure 7: Nominal question examples*



**Ordinal questions** have a list of categorical answers or choices that are ranked according to their values and order. By using ordinal questions, researchers aim to ask respondents to choose their preferred responses from a list of numerically ordered options. Ordinal questions are used for measuring respondents’ knowledge, preferences, or behavior concerning a variable’s ranked

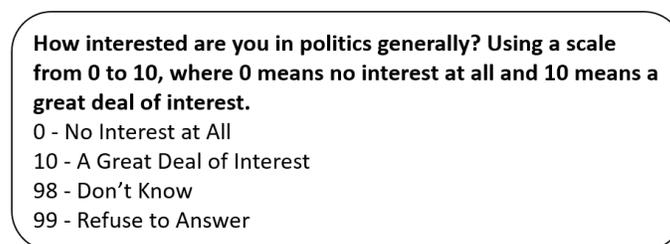
values. However, given the fact that they do not measure values at equal intervals, ordinal questions should be used as categorical rather than scaled variables. Ordinal questions provide the most reliable *median* of responses and are usually used as both independent and dependent variables in bivariate analysis. Two examples of ordinal variables are as follows (Ibrahimi 2020).

Figure 8: Ordinal question examples



**Interval questions** refer to the type of questions that include categorical responses that are ranked and there is an equal interval between each category. For example, the distance between 20 and 30 years old is the same amount of time as the distance between 60 and 70 years old. Interval questions are the commonly used questions in survey research. These questions usually measure respondents' feelings, levels of satisfaction, the strength of agreement, likelihood, and the like in interval scales. They are regularly used for obtaining the *mean* of responses in univariate analysis and are the most commonly used as dependent variables in bivariate and multivariate analysis (Haan, & Godley 2017; Great Brook). The following is an example of interval variables (Canadian Election Study 2015).

Figure 9: Interval question examples



For some examples of the questionnaire, questionnaire design, and characteristics of survey question see the following documents: Canadian Election Study 2015; Akseer et al. (2019); Ibrahimi (2020); Adib & McDevitt (2018).

#### When and when not to use this particular method

Surveys are a useful method for studying a large sample from which a research expects to generalize about a target population's knowledge, attitudes, or behavior concerning a particular issue or policy. Surveys are also a good method of inquiry for "gathering demographic data that describes the composition of the sample" (McIntyre, 1999, p. 74). For example, this method helps know the gender and ethnic dimensions of a population or the income or education categories in a population and their impact on research variables.

Furthermore, it is a useful method for collecting original or firsthand data from a large and diverse population when the purpose of research is generalization or provision of a general image about a target population's views on a particular problem, process, or policy. Surveys provide policy-relevant data that can influence policy-making or government decisions about a problem, process, or existing policy. The finding helps governments, policy-makers, and international organizations to adjust their policies and visions with realities on the ground and the target population's expectations and preferences. Moreover, survey research is used for producing data for building or updating national and global databases on certain issues. Those databases provide a significant basis for extended inquiries and are important sources of information about significant areas such as conflict, governance, state-building, etc.

There are two major conditions when the application of survey as a method of social inquiry is not suggested. First, surveys are a method of studying the relationship among a set of variables from a sample's view. The sample's views vary for different reasons and the researcher does not have control over those reasons. Therefore, the researcher is forced to see the result only from the perspective of the sample respondents which, in most cases, prevents an in-depth inquiry of the relationship between the research variables. Second, surveys should not be used when the study of a phenomenon requires in-depth knowledge and experience about the historical or cultural context of that phenomenon (Pinsonneault & Kraemer 2015). Application of surveys in such a condition requires intensive training of the researcher about the context of the inquiry, which in most cases is not possible. In such situations, it would be preferable to study the phenomenon by using qualitative methods such ethnography or observation.

Finally, the study of a large sample is difficult at best in conflict zones or among war-affected populations. The difficulty of access to all required demographic segments of a sample could lead to sample bias and the random or systematic exclusion of particular communities. This condition affects the representativeness of the sample, and therefore, the accuracy and validity of the data. Therefore, in the absence of a clear strategy that guarantees the inclusiveness of the research sample, a survey is not suggested as a method of inquiry. The types of data collection strategies

such as snowball sampling or purposeful sampling that are regularly used in conflict zones are also not useful for survey research.

#### Best practices for policy relevant research (the how to)

When survey research is inspired by policy gaps and needs, the quality of a survey is not only evaluated by its sample size or topics but also by the amount of attention the survey results receive in the policy arena or the public. Timely surveys that deal with important and ongoing problems produce the best policy-relevant results and can receive broader attention. Such results inform policy-makers with ideas and recommendations to adjust or revise policy and programs in response to ongoing demands, emerging problems, and changing preferences and behaviors. Survey results, in this context, provide the basis for governments and other sectors to create new policies or trends in accordance with their citizens' or clients' expectations, preferences, and concerns.

International attention and investment in increasing awareness about peacebuilding, good governance, development, and foreign aid in countries like Afghanistan have motivated survey research at national, subnational, and local levels. Recent surveys in the past three years in Afghanistan were motivated and directed by such demands for policy-relevant research. They have been timely and have received broad attention from the government, the public, the media, and other relevant sectors. These studies helped a variety of internal and external stakeholders to learn from policy gaps and unfavorable experiences and develop alternative measures for better policy practices. Studies were also motivated by a national and international demand for increasing awareness about the causes and conditions of shortages, slow processes, and deadlocks of development, state-building, and peacebuilding in Afghanistan. Besides touching upon those causes and conditions, recent surveys have introduced methodologies and concepts that can facilitate further research and follow-ups in certain areas of governance, conflict management, and development in the future.

For comprehensive results, some policy-relevant surveys use large sample data from previous national and/or international studies and databases, thus benefiting from alternative sources beyond the firsthand data. Having access to an existing database helps understand what information is needed, what is the best method to acquire the information, how survey data alone is not sufficient for providing the information, and how survey data can be supported and complemented by using non-survey data and existing databases (ASA 1995). A great example is the Global Conflict Risk Index which examines political, security, economic, social, and geographic variables in conflict contexts by using data from existing databases (De Groot et al. 2015). Policy-relevant surveys in Afghanistan can be applied more effectively and objectively by using such databases that are based on large, comparative, and evidence-based data (see CIFP).

#### **Analyzing and Interpreting results**

The interpretation or analysis of survey data requires statistical methods usually by using SPSS or STATA software. These two software products help conduct univariate, bivariate, and multivariate analyses of the data. (Haan & Godley 2017). Univariate analysis provides results about responses on a single variable. For example, the average income of respondents or their

level of interest about a political party or a policy is obtained by univariate analysis. Univariate analysis produces such statistics as frequency tables and measures of central tendency including the mean, median, and mode of responses. For example, Table 6 presents the frequency of responses about the level of respondents' support for the post-2001 political system in Afghanistan (Ibrahimi 2020) and Table 7 indicates the mean, median and mode of Canadians' responses concerning their interest in the 2015 Federal election (Canadian Election Study 2015).

*Table 8: Respondents' support for the post-2001 Afghanistan political system*

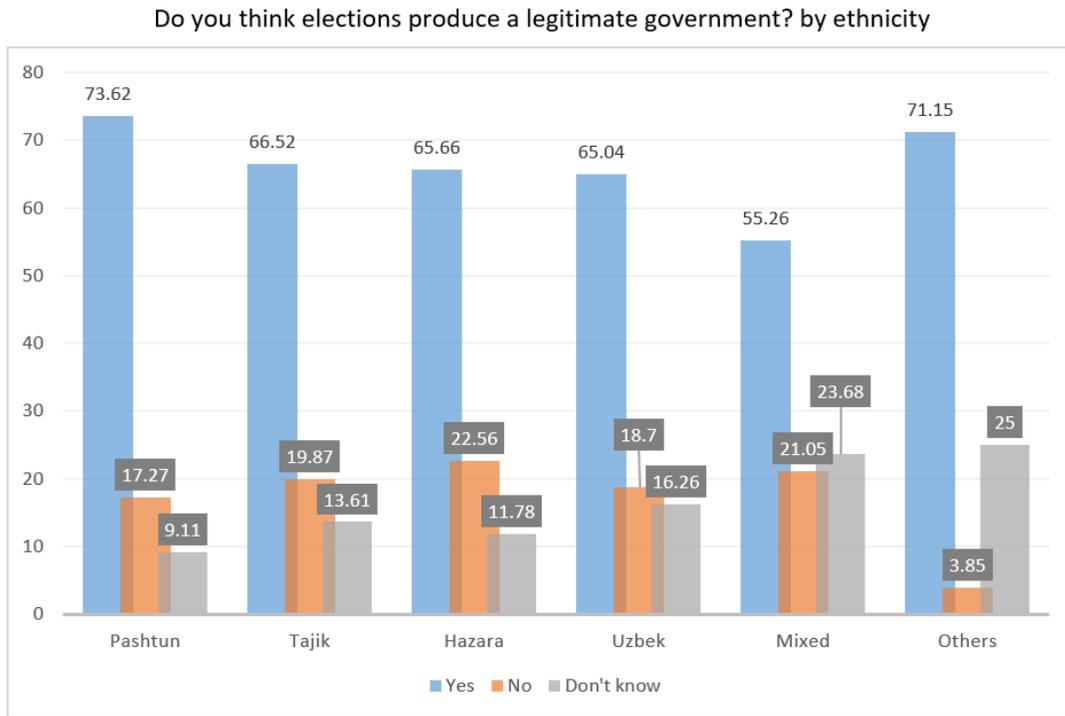
Responses	Percent age
Strongly support	40.34
Somewhat support	27.6
Somewhat oppose	20.53
Strongly oppose	9
Don't know/Undeclared	2.53

*Table 9: Canadians' responses towards federal elections*

Statistics		
Interest in Federal Election		
N	Valid	11058
	Missing	556
Mean		6.96
Median		7.00
Mode		10
Std. Deviation		2.652

By contrast, bivariate analysis produces data about respondents' views on the relationship between two variables. For example, the relationship between respondents' education and their level of support for democracy or the relationship between respondents' ethnicity and their level of agreement with a regime type, etc. The following figure is an example of bivariate analysis of survey data which shows that ethnic affiliation made little difference to respondents' views about elections as a mechanism of political legitimation in Afghanistan (Ibrahimi 2020).

Figure 10: Bivariate analysis of survey data



Finally, multivariate regression analysis allows researchers to control for a third or fourth variable while you are testing the relationship between two variables. For example, when you are interested in knowing the relationship between the level of education and the level of support for democracy, multivariate analysis allows you to also control for a third variable (i.e. ethnicity) and examine whether the relationship between education and democracy varies by ethnicity or not. Likewise, when you aim to test the relationship between respondents’ province of residence and their level of support for government, multivariate analysis allows you to examine how other factors such as respondents’ level of education, their level of income, ethnicity, political background and the like influence their level of support for the central government. As such, you can produce a comprehensive statistical model through multivariate analysis of survey results that not only analyzes the relationship between two variables but also indicates the role of a context or other conditions on respondents’ views (Haan & Godley 2017).

#### Examples and relevant application

Most recent examples of survey research in Afghanistan include Akseer et al. (2019); Ibrahimi (2020); Adib & McDevitt (2018). Each of these surveys has implications in specific areas. The first survey, which is a broad study of the Afghan people by the Asia Foundation, covers people’s views on various aspects of political, economic, security, and social issues and events in the country. The two latter surveys concentrate on specific areas including conflict resolution and corruption respectively. The survey on conflict resolution examines people’s views about mechanisms of conflict management and the political system in Afghanistan and has significant implications in

Afghanistan's ongoing peace process. The survey by Integrity Watch Afghanistan: *The National Corruption Survey*, studies the perceptions and experiences of corruption in Afghanistan from people's perspective and has significant implications for political development, particularly work towards state reforms and accountability.

## Focus Groups

### Overview

Focus group research is a qualitative method of research in which the researcher and his/her research team bring together a small number of individuals and discuss with them the topic of study. Among other qualitative methods, focus group methodology is broadly compared with individual interviews that produce deep and detailed information from individual participants. By contrast, focus group research relies on the variety of perspectives and experiences that participants reveal during their discussion on different aspects of the research topic. As Morgan & Hoffman (2018) puts it "focus groups are especially useful for investigating the extent of both consensus and diversity among the participants, as they engage in sharing and comparing among themselves with the moderator in a facilitating role." In general, focus groups are used as a research method when the researcher identifies that the type of information produced by group discussions or multilateral conversations provides the most appropriate empirical data for studying the topic of interest. Focus groups research is generally used for exploratory and context-specific research when the discussion and analysis benefit from social dynamics and collective reflections on a particular topic of interest (Berdahl & Archer 2015; Halperin & Heath 2017).

### Key Words

Focus Group, Moderation, Context, Social Dynamics, Content Analysis

### Concept Definitions and Terms

Focus group research is a type of group conversation for collecting data on topics that are context-specific and include social dynamics. Participants in a focus group not only share their views but also the social or community views about certain questions or issues. Focus group participants are selected purposefully, not randomly, based on their level of experience with or level of knowledge and information about the research topic.

In focus group research, a series of group discussions are designed by the researcher and led by a moderator. In focus group research, unlike surveys, the researcher or moderator does not use a pre-prepared questionnaire or ask questions with pre-determined answers. They provide a few open-ended questions to guide or motivate participants to focus on discussing the topics rather than simply answering the questions (Halperin & Heath 2017). The expectation of using this method is that relevant information will emerge from the group discussions. It is also expected that the discussions might take different directions and create new questions that will motivate the group to talk further.

Focus groups are organized in a structure where the moderator and researcher mostly observe the discussions, and sometimes they motivate the participants to stay within the focus and scope of the research. The researcher and moderator do not usually take notes, rather a notetaker is recruited to observe and record what is said (Berdahl & Archer 2015; Halperin & Heath 2017).

Since focus group research is highly exploratory, the researcher provides a few open-ended general questions around the themes of research to facilitate focus group discussions rather than limiting the respondents to provide answers to particular questions. Group discussions regularly begin with a brief introductory statement about the research topic and the dimensions and expectations of the research, rather than a question, and proceed with guiding questions or statements that lead to intense discussions. At the end of the focus group, some researchers use a short complementary questionnaire that includes demographic questions and key questions related to the central theme of research. Using such questionnaires is useful for putting the group discussion into context and facilitates the interpretation of results (Carbert 2003).

### **Focus Group Structure**

This part addresses three important questions about the structure of focus group research: who is researched? Who researches? And how the focus group research is organized.

An important part of the focus group structure is the size and number of focus groups. The size of a focus group and the number of groups depend on several important issues including the homogeneity or heterogeneity of the target populations, the level of research topic's complexity, and the purpose of research. Taking these issues into consideration, focus group research usually requires 10-20 groups of participants. The number of groups depend on the size of geographic area or the demographic characteristics of a population that a researcher wants to cover in a research.

Louise Carbert, for example, used 14 focus groups of women participants in her research on women's political participation in the Atlantic provinces of Canada. The size of each group also differs based on the same issues. For example, a larger and more heterogeneous population and the study of a more complex topic require bigger groups and vice versa. In general, every group typically requires 5 to 10 participants, although some authors recommend 6 to 12 participants (see Morgan & Hoffman 2018; Berdahl & Archer 2015).

The rationale for the size of focus group also stems from the goal of focus group research that is every group should yield diversity and yet they should not include too many participants that would make the sharing of thoughts, beliefs, and experiences difficult. Therefore, the composition of a focus group (or who participates) is more important than the size of groups (Berdahl & Archer 2015).

Another important aspect of the structure of focus groups relates to the role of researcher(s) and the organization of groups. Focus group research includes a team of data collectors under the supervision of a researcher. Therefore, besides the researcher who mostly observes the

discussions and the way data is produced, in every focus group, there is a moderator who leads discussions and motivates participants to stay in the scope and direction of the research and sometimes ask questions to avoid the progression of irrelevant talks. The notetaker records the conversations and responses. Besides notetaking, some researchers prefer to audio record the discussion which requires the permission of all participants.

Every focus group begins with a brief introduction by the moderator about the topic and purpose of research, key themes, and questions that exist about the topic, and the rule and duration of focus group discussion. Focus group duration is typically one to two hours (Berdahl & Archer 2015; Halperin & Heath 2017).

Every focus group must be designed in a way that avoids the authority or domination of particular individuals during the discussion. For example, the inclusion of teachers and their students, or employer and their employees in focus groups on topics related to teaching or workplace issues, respectively, might lead to the hiding of some parts of the information by participants. Therefore, it is important to look at the focus group combination and its relations with the topic of study very carefully and avoid situations where power dynamics might affect the discussion in a focus group, such as the inclusion of teachers and students or employers and employees as participants in a focus group on topics related to teaching or workplace affairs, etc.

The composition of focus groups, if it is not organized carefully, could cause harm, breach trust, and lead to the emergence of inaccurate information. Therefore, sample selection and group combination are important elements of focus group research. Unlike survey, focus group participants are selected purposefully, rather than randomly. This means that they are selected based on their level of knowledge about and level of involvement in or experience with the topic of research. Therefore, the respondents' relevance to the topic must be considered as a key element of focus group research.

It is also important to consider the sensitivity of the research topic and its relationship with individuals that are selected for focus group research. For example, when individuals in ethnically or racially divided societies are selected to discuss a topic that has group dynamics and sensitivities, participants may hide or change facts during group discussions or polarize the focus group which makes a comprehensive data collection difficult (Berdahl & Archer 2015; Halperin & Heath 2017).

### **Analyzing and Presenting Focus Group Results**

Focus group data can be subjective and difficult to interpret. Therefore, certain methods are required to carefully conduct rigorous data analysis. Different methods of data analysis such as content analysis and interpretive content analysis are typically used for analyzing focus groups and other types of qualitative data (see Drisco & Maschi 2015; Elliott & Timulak 2005; Morgan & Hoffman 2018). The application of any data analysis method depends on the scope and purpose of research and the type of information that a researcher aims to publish.

This manual provides a guide to content analysis as a systematic method of analyzing focus group data. The method can be applied for analyzing other types of qualitative data such as individual interview transcripts, participant observation field notes, etc. A content analysis of focus group data requires a variety of tasks including data preparation, data coding, data analysis, and presentation of findings.

The first step of content analysis includes the preparation and review of the data which is obtained as written records or notes from group discussions. These notes or transcripts become the main source of data analysis (Stewart et al. 2007; Onwuegbuzie et al 2009).

After data preparation, coding is used for a systematic description and analysis of the data. Coding, in this sense, is described “as the *critical link* between data collection and their explanation of meaning” (Charmaz 2001; Schreier 2014). It is important to note that the coding processes can significantly affects the quality of focus groups’ results. Therefore, it is important to take the coding steps carefully to make sure that during the process of reducing the data into readable codes and categories parts of the data are not eliminated or ignored.

The coding process should follow three general steps including coding, categorizing, and theorizing (Saldana 2013). These three steps help researchers to move from reviewing the raw data towards producing specific results and conclusions. The three key steps of coding, as articulated in Saldana (2013), are described below:

1. **Coding:** At this very preliminary stage, researchers review the data and identify codes that are relevant to the research question. A code in qualitative data analysis can be a word or a short phrase in the data that are relevant to the research question or related to the research interest. In general, there are three types of codes that a researcher can identify and list in the data including *descriptive codes*, *In Vivo codes*, and *eclectic codes*.

Descriptive codes are simply used for summarizing topics of interest in the data. For creating a descriptive code, researchers identify a statement of interest in the data and then summarize it in a single word. This word can be used as an indicator or factor in data analysis. In the following example (Figure 9), the capitalized word in the right column indicates a descriptive code.

Figure 11: Descriptive code

<p><sup>1</sup> I notice that the grand majority of homes have chain link fences in front of them. There are many dogs (mostly German shepherds) with signs on fences that say "Beware of the Dog."</p>	<p><sup>1</sup> SECURITY</p>
---	------------------------------

(Source: Saldana, 2013)

By contrast, *In Vivo* codes capture several options or preferences in a statement within the data. The following example (Figure 10) on students' opinion about their mentor indicates three *In Vivo* codes. The capitalized words in the right column are *In Vivo* codes.

Figure 12: In Vivo codes

<p><sup>1</sup> He cares about me. He has never told me but he does.</p>	<p><sup>1</sup> SENSE OF SELF-WORTH</p>
<p><sup>2</sup> He's always been there for me, even when my parents were not. He's one of the few things that I hold as a constant in my life. So it's nice.</p>	<p><sup>2</sup> STABILITY</p>
<p><sup>3</sup> I really feel comfortable around him.</p>	<p><sup>3</sup> "COMFORTABLE"</p>

(Source: Saldana, 2013)

Finally, *eclectic* codes are not specific codes, but they include "first impression" phrases in the data that have emerged from the perspectives of participants and their experiences. These codes emerge when researchers rely on a particular "passage of data to decipher its core meaning" (Saldana 2013). These codes are important because they introduce patterns. The following example (Figure 11) on a mother's statement about her "teenage son's troubled school years" indicate five eclectic codes (Saldana 2013). The capitalized words in the right column of Figure 11 are eclectic codes.

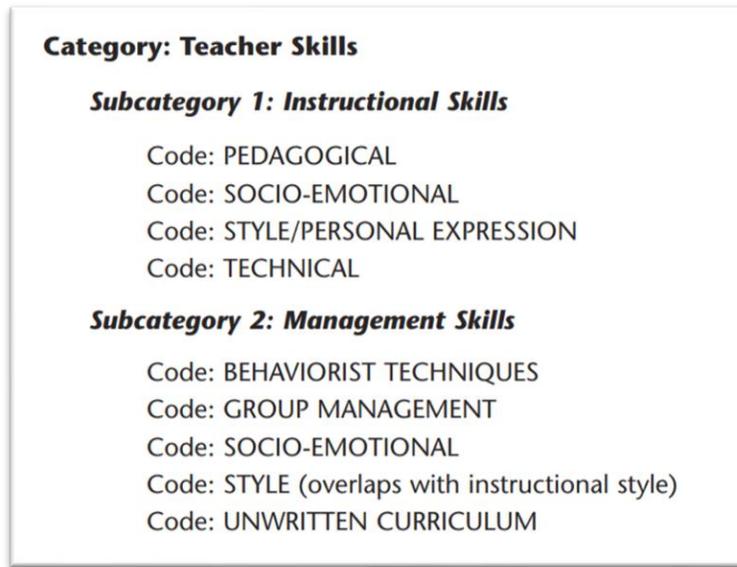
Figure 13: Eclectic codes

<sup>1</sup> My son, Barry, went through a really tough time about, <sup>1</sup> MIDDLE-SCHOOL  
probably started the end of fifth grade and went into HELL  
sixth grade. <sup>2</sup> When he was growing up young in school <sup>2</sup> TEACHER'S PET  
he was a people-pleaser and his teachers loved him to  
death. <sup>3</sup> Two boys in particular that he chose to try to <sup>3</sup> BAD INFLUENCES  
emulate, wouldn't, were not very good for him. <sup>4</sup> They <sup>4</sup> TWEEN ANGST  
were very critical of him, they put him down all the  
time, and he kind of just took that and really kind of  
internalized it, I think, for a long time. <sup>5</sup> In that time <sup>5</sup> THE LOST BOY  
period, in the fifth grade, early sixth grade, they really  
just kind of shunned him all together, and so his  
network as he knew it was gone.

(Source: Saldana, 2013)

- 2. *Categorizing or codifying:*** At this stage of data analysis, researchers create categories for relevant codes that are identified and listed in the earlier coding stage. Categorization of codes is a process that permits data to be “segregated, grouped, regrouped and relinked in order to consolidate meaning and explanation” (Grbich, 2007, 21; Saldana 2013). It also helps create more abstract and presentable concepts and meanings. Figure 12 presents an example of categorizing codes. This example explains how a major category (i.e., Instructional Skills) emerge from two relevant subcategories that, in turn, are formed of relevant codes. This example, first, shows how codes can create subcategories and, next, how subcategories can form a more abstract major category. Both major and *sub* categories are useful for clarifying and presenting main issues in the data and their empirical bases.

Figure 14: From Codes to Category

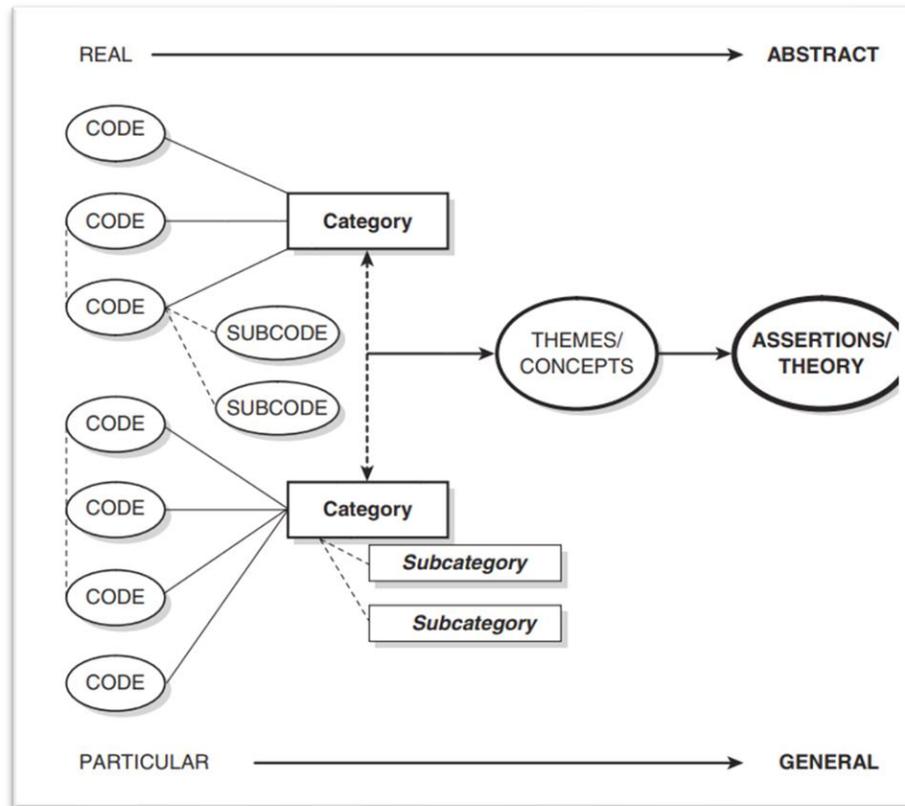


(Source: (Lewins & Silver, 2007; Saldana 2013)

The above example can be defined as a first cycle categorization of codes in the data. However, the analysis of qualitative data requires further scrutiny at this stage. As mentioned in the beginning of this section, qualitative data is mostly subjective and not easy to interpret. Therefore, researchers need to review the data at least a second time by using the same method of coding and categorizing. This method is helpful for identifying further facts, codes, and themes in the data which help develop existing categories and create new subdivisions for a rigorous analysis (Saldana 2013).

- 3. Theorizing or moving from data to results:** The categorization of codes helps organize and shape the diverse and raw data into classes of more general themes and concepts. However, these concepts are not presentable as the final result of the research (Richards and Morse 2007). Therefore, researchers are required to take a third step which is known as theorization of assertion. At this stage, they will need to explain the systematic relationship and interrelations between concepts and themes that have emerged from the earlier stages of data analysis (Corbin & Strauss, 2008). Theorization allows for the articulation and explanation of the meanings of concepts/themes and helps provide a clear and presentable result. Although this stage of data analysis is identified as theorization by some scholars, the result does not need to always develop a new theory in its conventional sense (Saldana 2013). The most important purpose of data analysis at this stage is to provide a meaningful explanation about the relationship between concepts that are created through coding and categorizing and their link with the research question or the problem of interest which will lead to clear results/findings. Figure 13 illustrates the process of producing results from codes and categories.

Figure 15: From codes to results/theory



(Source: Saldana, 2013)

Focus group findings are usually presented as qualitative statements. In this case, key findings or assertions are listed and then supported by empirical data which are usually codes and categories that have emerged from the data analysis processes. Those findings can also be presented numerically and statistically in the form of frequencies, percentages or inferential statistics (Morgan & Hoffman 2018; Schreier 2014).

#### When and when not to use this particular method

Focus group research is a very useful method of research for studying context or community perspective about a social issue, policy, process, etc. Applying the focus group method is based on the assumption that the type of data needed for studying a topic with community or group dynamics can be better produced through group interviews or group conversations than individual interviews or surveys. It is also used for exploratory research expecting that the knowledge about a problem emerges from group discussions rather than individual interviews or surveys.

Some researchers also use focus groups for creating survey questions when sufficient literature about the topic does not exist or when he/she wants to explore more complex aspects of a problem from the perspective of a relevant population or community. Focus group research is

also used for producing accurate information about what people think and when community dynamics and knowledge or group processes are important for studying a problem. In general, the focus groups method allows a researcher to dig beneath the surface of opinion that is captured by a large survey or other forms of research and helps explore details, accuracy, and nuance of the existing knowledge about an issue, policy, or process (Berdahl & Archer 2015).

Generalization is rarely a goal of focus groups, especially given the exploratory nature of this method. Moreover, focus groups should not be used for studying topics that are sensitive for group discussions, such as ethnic or racial issues in highly divided societies or communities. In such conditions, participants may hide or change facts during group discussions or polarize the focus group which makes comprehensive data collection difficult, if not impossible (Berdahl & Archer 2015; Halperin & Heath 2017).

Finally, focus group research is applicable when it is possible to access participants with knowledge about the topic who are interested and willing to communicate about all aspects and dynamics of the research topic. Moreover, the application of focus group research requires that the topic is understandable to group members, and moderators are highly skilled in managing group conversations. By contrast, this method is not useful when participants are deeply divided around aspects of the topic and when polarization can emerge during the discussion. In such situations, trust cannot be established among participants to present their real ideas about the focus group topic(s). In such an unfavorable environment, participants will censor their statements or conform to the dominant group opinions, decreasing the validity of data (Carey & Asbury 2012; Berdahl & Archer 2015). In general, the application of focus group research is suggested for studying homogeneous groups rather than heterogeneous groups (Berdahl & Archer 2015).

Best practices for policy relevant research (the how to)

Focus groups shed light on gray and complex areas of policy that are missed by other structured methods of research such as surveys or polls (Gupta 2001). Moreover, focus groups can provide a comprehensive understanding of social attitudes concerning a policy, instead of measuring them (Luntz 1994). For example, focus group findings explain how and why a community responds or reacts to policy, instead of counting the frequency of responses on that policy. This can help policymakers gain an original picture of the form and shape of reaction to a policy, rather than examining a policy merely through numbers. Policy-makers who are informed of community dynamics and social contexts of certain policies or programs frequently ask questions about the details of what a target community thinks about it. Focus groups are a useful method for providing answers to such questions.

Moreover, focus group research can be useful when a community perspective is needed for creating or revising policies. It is useful for developing new policies based on a community's perception of a problem or for revising existing policies in the light of the community's experiences and expectations.

## Examples and relevant application

Focus group research is used for both exploring in-depth and context-specific information and providing the basis for the survey questionnaire. For example, Louise Carbert (2003) studied women's political participation in the Atlantic provinces of Canada by using a series of focus groups. This research is based on findings from 14 focus groups involving 126 rural women leaders in the four Atlantic provinces of Canada. The research focuses on "the dominance of regional economic development programs in local affairs, and the associated consequences for the electoral ambitions of women in leadership positions." The result provides a detailed and in-depth image of why and why not rural women participate in politics in these four provinces.

By contrast, Luntz (1994), uses focus group research to evaluate the accuracy of findings of survey and polling. Luntz believes that survey and polling provide information on how people think, but focus groups inform us about how they feel/think about what they think. He believes that quantitative research only covers respondents' "yes or no," "agree or disagree" and similar answers but fails to explain what they really think about providing those answers or filling survey questionnaires. Focus group research is necessary for providing answers to those questions. Therefore, any topic that is broadly researched, needs qualitative data to test its accuracy and provide information about community dynamics (Luntz 1994).

Finally, depending on the topic of interest, focus groups can be useful for studying sensitive issues. However, careful judgment is needed about studying sensitive topics through focus groups as they are not appropriate if people feel hesitant to discuss the particular topic in public. Community-based mental health issues are topics that can be studied by using focus group research depending on participants' level awareness and services available. The application of this method helps examine both community dynamics and context-specific aspects of mental health problems. For example, research by *Link Health* in 2016 uses focus group research to study mental health awareness, access to mental health, and mental health related stigma in the Afghan community of Victoria, Australia (Karimi 2016). This research provides a community perspective about mental health issues among Afghan residents of Victoria, aiming to increase community awareness and develop consulting services in the target area based on the community's perception of and experience with the problem (Karimi 2016).

# Part 3: Research and Evaluation Presentation

## Policy Briefs and Policy Papers

### Overview

This section will cover the fundamentals of policy-oriented report writing. A chart will clarify the use-case for a variety of policy reports and a toolkit will cover the conceptual and stylistic elements of report-writing. The structure of a policy paper will be described in detail, as well as the formation of scenarios, recommendations, and policy options. This toolkit will also feature innovative tools to enhance the report-writing process.

### Concepts definitions, key words and terms

**Policy-oriented report writing** is ‘concise, specific, objective, persuasive, and focused on practical decision making’ (McIvor 2018). The target audience is the policymaker and by extension, stakeholders including civil society organizations, lobbyists, activists, the media, and the private sector. Reports are tailored to this audience and thus solution-oriented, short with essential information, in plain language, and visually appealing.

**Academic writing** on the other hand is exploratory and theoretical, focused more on knowledge base expansion than on the advancement of pragmatic solutions. The target audience is often other academics with advanced knowledge of the topic (McIvor 2018). Reports tend to follow a strict format and are peer-reviewed, longer and more verbose.

Examples of policy-oriented reports include:

- **Policy Briefs:** Concise stand-alone reports (1-10 pages) that outline an issue, provide a balanced, evidence-based analysis of policy responses, and offer a recommendation. The audience includes decision-makers with little time to read comprehensive documents.
- **Policy Papers or Policy Studies:** Longer, rigorous reports that use original research and draw on existing literature to provide a policy recommendation. The methodology is outlined and findings are situated in the existing literature. The author weighs the risks and opportunities of different policy options, and makes their recommendation based on their analysis. This might be accompanied by a detailed implementation plan as an annex.
- **Rapid Evidence Reviews.** Rapid reviews of existing evidence that use systematic methods (clear methodology, inclusion/exclusion criteria, limited appraisal of sources) but limit the breadth or depth of the process due to time constraints (Grant and Booth 2009). Usually completed in under 5 weeks, depending on available resources and expertise.
- **Early Warning Brief.** Forward-looking document that can include best, worst, and most likely scenarios based off indicator trend analyses, key events, and stakeholder analyses.
- **Baseline Assessments.** Studies conducted at the start of a project, programme, or policy to facilitate the assessment of an intervention over time. This report will draw on an evaluation framework with specific quantitative and/or qualitative indicators.

Table 10: Policy documents and their use cases

Type of Document	Use-case	Average Word Count	Intended Outcome	Essential Elements	Example
Policy Brief	Timely, concise information for an upcoming decision	Under 5000 words	Issue-specific policy recommendations	Introduction, Background, Issues, Recommendations	<a href="#">IWA 2017 Policy Brief</a> (Afzali and Timory 2017)
Policy Paper	Time allows for a comprehensive analysis of an issue and policy options	Around 15000	Policy recommendations grounded in literature and with some original research	Executive Summary, Context and Importance of issue, Methodology, Literature Review, Analysis of Findings, Policy Options and Recommendations, Implementation, Conclusion, Appendices, References	<a href="#">IWA Policy Paper</a> (“Senior Appointments and Corruption within Kabul City Police” 2015)
Early Warning Brief	Foresight on critical risks/threats	No clear standard	Best, worst, and most likely scenarios based on existing evidence and stakeholder interests/positions	Executive Summary, Context, Indicator trends, Stakeholder analysis, Scenarios	<a href="#">Conflict Diagnostic</a> (Boucher et al. 2020)
Rapid Evidence Review	More rigorous than a literature review but less exhaustive and more timely than a systematic review	No clear standard	Systematic yet rapid appraisal of evidence base	Executive Summary, Background, Methods, Findings, Discussion, Limitations, Conclusion	<a href="#">Public Procurement Reform – Rapid Evidence Review</a> (Leacock 2013)
Baseline Assessment	Outset of a new policy, program, or project	No clear standard	Starting point for evidence-base and foundation for impact assessment	Executive Summary, Context, Main objectives and expected impact,	<a href="#">UN WOMEN Baseline Assessment</a> (“The Joint Program on Essential

				Approach (scope and methodology), Analysis and Reporting, Ethical considerations, Findings, Limitations, Conclusion and Recommendations, Annexes	Services Package (Esp) for Women and Girls Subject to Violence in Vietnam: Baseline Assessment Report 2018” 2019)
--	--	--	--	--	---

Best practices for policy-relevant report writing (the how to)

**Policy Paper Outline:**

- Title**
- List of abbreviations**
- Table of Contents**
- Executive Summary**
- Introduction**
- Methodology and Limitations**
- Analysis of the Policy Issue**
- Policy Options**
- Recommendations and Conclusion**
- Appendices and Annexes**
- Bibliography**

**Title**

A strong title will set your paper apart and likely increase your readership. Make sure your title clearly describes the substance of the paper without being too long or uninteresting. It is important here to consider the target audience, as a more technical title will appeal to an audience of specialists, but a concise and interesting title will draw in decisionmakers with little time.

See the following examples:

- 1) Participatory lived experience research: Barriers and enablers for social inclusion for people with psychosocial disability, in Afghanistan (Allan et al. 2018).

Notice the use of the colon to separate two parts of the title. The first section identifies the methodology, while the second section identifies the location (Afghanistan), research subject (people with psychosocial disabilities), and type of findings (barriers and enablers) to expect. While a bit long, this title is descriptive and would likely appeal to research specialists seeking some combination of the elements described by the title.

- 2) An equity analysis of utilization of health services in Afghanistan using a national household survey (Kim et al. 2016).

This title is more concise the first, while still identifying the methodology (equity analysis), topic (utilization of health services), location (Afghanistan), and in this case, the source of the analysis (national household survey). This was written for an audience of health equity specialists or academics.

- 3) Exploited by whom? An alternative perspective on humanitarian assistance to Afghan women (Barakat and Wardell 2002).

This title starts with a rhetorical question and provides a hook for an ‘alternative perspective’. Those without an understanding of the subtext in this title would likely miss the appeal of it. The author is speaking to an audience that is already familiar with the dominant narrative regarding humanitarian assistance for Afghan women, is likely jaded by the situation, and would be interested in alternative perspectives to explain the challenges they face. The article sacrifices a clear description of what will be described for a more interesting and concise title. The audience here might include humanitarian actors, policymakers, and women’s organizations.

## List of Abbreviations

In longer, more technical policy papers, it is often a good idea to use a list of abbreviations when there are many acronyms for organizations, programmes, and issues. If the article repeatedly uses non-English words, a currency conversion at a specific time, or a term that would not make sense without context, consider adding them to the list of abbreviations. It is important to only add a list of abbreviations if necessary and to only add terms that are used repeatedly.

The list of abbreviations will feature in the beginning of a document after the table of contents and will be in alphabetical order.

Take a look at the following example from the [2016 IWA National Corruption Survey](#) (Mackenzie 2016):

### ACRONYMS

Afs	Afghani (Afghan unit of currency)
ANA	Afghan National Army
ANP	Afghan National Police
ANSF	Afghan National Security Forces

BBC	British Broadcasting Corporation
CDC	Community Development Council
Chief	Executive Officer (post created for Abdullah Abdullah in the National Unity Government)
CSIS	Centre for Strategic and International Studies
CSO	Central Statistics Organization
GDP	Gross Domestic Product
GIRoA	Government of the Islamic Republic of Afghanistan
HOO	High Office of Oversight
ISIS	Islamic State in Iraq and Syria, also known as Da’esh, IS (Islamic State) and ISIL (Islamic State in Iraq and the Levant)
IWA	Integrity Watch Afghanistan
MRRD	Ministry of Rural Rehabilitation and Development
NCS	National Corruption Survey (biennial publication of IWA)
NDS	National Directorate of Security
NRVA	National Risk and Vulnerability Assessment
NSP	National Solidarity Program
NUG	National Unity Government: Name given to the US-brokered power-sharing agreement between Presidential candidates Ashraf Ghani and Abdullah Abdullah, in the wake of a hotly contested election in 2014
OS	Original Sample (for the figures)
PC	Provincial Council
SIGAR	Special Inspector General for Afghanistan Reconstruction Body created by the US Congress to provide independent and objective oversight of Afghanistan reconstruction projects and activities.
TAF	The Asia Foundation
UNODC	UN Office on Drugs and Crime
	1 USD = 68 Afs (average used in the data, with source)

## Table of Contents

The table of contents outlines the structure of the paper and provides the reader with a navigational menu of clickable headings and subheadings when reading online. An effective table of contents makes use of indentation and decimal numbering to give the reader visual, intuitive cues about the structure of the paper.

See the example below of a shortened table of contents:

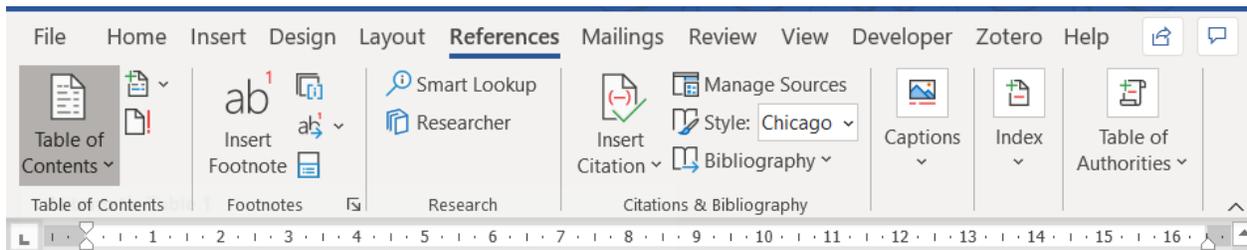
### TABLE OF CONTENTS

ABOUT INTEGRITY WATCH .....	i
ACKNOWLEDGEMENT .....	ii
ACRONYMS.....	iii
EXECUTIVE SUMMARY.....	viii
INTRODUCTION .....	1
1. GENERAL TRENDS .....	6
1.1. General Perceptions: Overall, Afghans appear optimistic about their home areas .....	6
1.2. Perception of Political Situation.....	9
1.3. Major Problems Facing Afghanistan .....	10
1.4. Addressing National Problems: Government, International Community and Taliban .....	15
1.5. Perceptions of Public and Civil Institutions in Society .....	17

1.6. International Community.....	18
1.7. Perceptions on Peace Talks and Reconciliation .....	19
1.8. SUMMARY OF PART ONE .....	21

If using Microsoft Word, the author can use the References tab to automatically create a table of contents that corresponds with the text in the headings and subheadings in the document. Note that Heading 1 in the “styles” section of the home tab will create a main heading in the table of contents, while Heading 2 will create an indented sub-heading. This table can be updated automatically with the click of a button, saving time when page numbers, headings, or sub-headings change in the document.

Figure 16: Table of contents in the Microsoft Word ribbon



The table of contents is often followed by a list of figures, tables, or maps. While mostly similar in format, the numbering system is slightly different from that of the table of contents. Each entry starts with “Figure”, “Table”, or “Map” with its corresponding number, either as a whole number (1,2,3) or as a decimal (1.1, 1.2, 1.3). See the example from the same IWA document referenced above (Mackenzie 2016):

TABLES

Table 1: In your opinion, do the people mentioned.....	17
Table 2: Total amount of bribes paid 2010-2016.....	35
Table 3: Effectiveness in fighting corruption.....	58

MAPS

Map 1: District Security Map 2016.....	2
Map 2: Level of satisfaction by province.....	8

**Executive Summary**

The executive summary asserts the significance of the issue, includes the most important findings, and concludes with recommendations. The executive summary is often completed last, when the author has completed all other sections. This is likely the only section to be read in its entirety by decision-makers and should be of priority for the author. The summary should aim to be a complete product on its own accord, avoid jargon as much as possible, flow logically and be written with a compelling voice. Be sure to avoid simply copying and pasting from other parts of the paper, as this will take away from the quality of the summary.

See the following example from IWA's [Senior Appointments and Corruption within the Afghan Mol: practices and perceptions](#) ("Senior Appointments and Corruption within Kabul City Police" 2015). Notice that the executive summary is kept to a single page, with only a paragraph describing the issue and seven clear recommendations. The policy issue is defined, this study's contribution is highlighted, and the rationale for changing the status quo is readily apparent.

## EXECUTIVE SUMMARY

Establishing meritocracy in appointments is an indispensable part of police reform and the establishment of the rule of law in post-conflict societies because the de-politicization of the police force and professional provision of security and law-enforcement depend on it. This study, however, demonstrates that meritocracy does not dominate the policy of appointments of the Mol and the ANP. Although the number of educated officials and police officers have increased and merit-based appointments are made to positions for which technical expertise is needed, connections and money are still instrumental in securing many senior positions within Mol and ANP. There also has been improvement in ethnic representation of the police chiefs, particularly in districts but to lesser extent in the Mol. Non-meritocratic appointments, nevertheless, has undermined police accountability and the relation between the ANP and the public. The on-going insurgency and the ANP engagement in fight against insurgents are also viewed to preclude the efforts to promote civilian policing. As the responsibility of providing security will lie on the shoulders of the Afghan security forces and international aid will dwindle after 2014, establishing meritocracy and fighting the corruption in the ANP appointment system seems more urgent than ever for long-term stability and durable peace in the country.

## Recommendations

- The new government needs to introduce and, more importantly, enforce clear educational qualifications and experience requirements for senior appointments and promotions.
- The leadership of the Ministry of Interior must be insulated from political influence and pressures from politicians for making appointments or promotions within the Mol and ANP.
- The HR office of Mol is responsible for overseeing appointments and promoting meritocracy. This study, however, shows that this office has had limited success in promoting merit-based appointments. The function of this institution and its procedures need to be reviewed and examined to ensure that it performs its envisioned function.
- There is currently no mechanism for obtaining public feedback about the ANP. An ombudsman office needs to be established, independent of Mol and ANP structures, for receiving public complaints and feedback regarding the police performance. The office must have the power to investigate the complaints and publicize them.
- Central to the corruption within Mol and ANP is the prevalent impunity and lack of transparency in dismissal of corrupt officials. To end the existing culture of impunity, the new government needs to take the prosecution of cases of corruption seriously and end secrecy surrounding the prosecution of cases of corruption and the decisions made on these cases.
- Reinstate the cyclical rotation procedure for all senior appointments at province and district level.
- While this report mainly deals with just perceptions of corruption, the fact that so many police officers are adamant that corruption is widespread within the Mol should warrant an in-depth investigation into the issue, as the legitimacy of the Afghan state is at stake.

## Introduction

The introduction is an opportunity to share important contextual information, provide essential definitions for the reader to understand the policy issue, explain the objectives behind the original research conducted, and to outline the remainder of the paper. Avoid sharing broader contextual information that, while important, is not essential to understand the specific issue. For example, if the policy paper is on meritocratic appointments in the Ministry of Interior or in the police department, there is no need to discuss at-length the history of policing in Afghanistan.

Take a look at the following example from IWA's [Senior Appointments and Corruption within the Afghan Mol: practices and perceptions](#) ("Senior Appointments and Corruption within Kabul City Police" 2015). Notice how, similar to the executive summary, the introduction starts with a general statement about police reform in post-conflict societies, but provides more information on the issue and how it plays out in Afghanistan. The intent of the research is clear, to apply pressure on policymakers and increase transparency and meritocracy for senior appointments. It goes on to introduce the study's research design and outlines the remainder of the paper.

## INTRODUCTION

Police reform is widely believed to be essential for long-term peacebuilding in post-conflict societies, since it is prerequisite for providing security and establishing the rule of law.<sup>1</sup> The three pillars of police reform are usually indicated as technical training, de-politicization of the police force and establishing a merit-based system of appointments and promotions.<sup>2</sup> This project aims at examining one of the key pillars of police reform in Afghanistan, that is, dynamics of senior appointments within the Ministry of Interior and the Afghanistan National Police. It looks at corruption and nepotism in senior appointments, which are inimical to the accountability of the police force and, consequently, to the rule of law in Afghanistan. The research project intends to put pressure on policy makers to take meritocracy and transparency in these appointments seriously by providing insights to key stakeholders and raising public awareness about this critical issue.

The project created a database to centralize the information collected through qualitative, interviews with approximately 250 individuals middle rank to senior police officials. The study collected limited quantitative and/or demographic data which will be included in the database and analysis. The study looked at the interviewees' past, qualifications and reputation, current affiliation and the reasons for which they were appointed. The sample is comprehensive enough to provide the very first study of such kind, allowing us to draw ideal-types of appointments and ideal-types of corruption related to them and what all this means for governance and state building in the context of transition.

The long-term impact of the research will hopefully be contributing to increased transparency and meritocracy for senior appointments within the ministry of interior and the police ranks, which would contribute to the success of police reform and the improved accountability of police force and the rule of law in Afghanistan.

The study is organised into two main sections, one dedicated to appointments and the other to the issue of corruption, which is linked to it. The section on appointments first discusses the data gathered through this survey, the perceptions of professional police officers with regard to the same appointments and finally the mechanisms presiding over appointments in the Mol. The perceptions of professional police officers are important because these men are supposed to be the protagonists of any successful effort to fully professionalise the Mol. This sub-section discusses topics such as the extent of professionalization so far and the various factors driving appointments. The concluding paragraph raises doubts about the suitability of a mere head count of professional officers as a measure of the meritocratic character of a system. The

section on corruption is divided in two sub-sections, one dedicated to evidence of anti-corruption efforts, as described by the interviewees, and the other to the description of corrupt practices, as provided by the same interviewees.

## **Methodology and Limitations**

Provide an outline of your methodology and include a discussion of limitations. The description of the methodology will depend on the type of research and it is useful here to draw on your research design outline. Provide specific information wherever possible, including for example numbers of interviews and interviewees, locations, time-periods of studies (longitudinal, cross-sectional, retrospective, etc.), and the type of study (observational, desk reviews, case studies). Provide a rationale for the type of study, for the selection of specific subjects or locations, and exclusion and inclusion criteria for documents.

The principle of reproducibility is key here and the policy recommendations will be more credible if the research process is as clear as possible. The description of the research needs to be sufficiently detailed so that the reader can see how the results were generated. This includes the discussion on limitations, where the author can outline selection biases, issues with validity and reliability, confounding variables, and necessary limitations imposed by ethics or security requirements. The researcher can also identify any mitigating strategies used to counter limitations.

Note the following shortened version of the methodology section in IWA's [Education Compromised? A Survey of Schools in 10 Provinces of Afghanistan](#) (Sharma and Afzali 2018). The sample size (276 schools), locations (10 provinces), time period (three years), variables of interest (operational status of schools, quality of infrastructure, access to sanitation, etc.) and selection criteria are clearly outlined. There is also a section on security assessments and limitations, which do well to explain omissions in the data due to the circumstances and the mitigating strategies employed by the researcher.

### **Methodology**

This research is based on primary data generated through field surveys and limited inspections of 276 schools in 10 provinces over a period of three years to assess the operation and physical condition of schools in Afghanistan. The objectives of the research were to assess the physical condition of school buildings and the ability of the facility to delivery education services. More specifically, operational status of schools, quality of school infrastructure, functionality of school support mechanism, access to sanitation facilities, availability of trained teachers, gendered information on student enrollment, quality and quantity of learning materials, and student-teacher ratio were assessed. The schools in the sample were built by specific donors in each sampled province and pre-selected for limited inspections. The sample was given to Integrity Watch by the donors to conduct its survey in accordance to the objectives set out prior to the design of the research.

The final sample used in the report included only those schools that are in the most accessible and secure areas of the provinces, making site visits possible for the surveyors. The sample selection criteria were 1) accessibility of the location and 2) the security situation of the surrounding area of the school. Generalizations made by drawing inferences from this sample survey can be applied to secure and

accessible areas in other provinces of Afghanistan. The problems highlighted in this survey may be even worse in more remote and less secure locations.

### 1. Security assessment

Prior to embarking on the field survey, a thorough security assessment of the districts in which schools located were to be surveyed was taken. This was done in consultation with the security department and the field team of Integrity Watch research department. Accessibility to each school facility was prepared based on a four-tiered, color-coded security ranking system: Green, Amber, Red, and Black. Schools surveyed were located in security zones classified as “Green” and “Amber.” In “Red” districts only schools located in the center of these districts were accessible, whereas in the case of districts identified as “Black” no schools were covered by field visits. Therefore, those schools were dropped from the survey. This includes 40 schools in Faryab funded by the NCA, 23 schools funded by CERP, and 121 schools funded by USAID in various provinces. Out of the schools selected, there were 10 schools in total that were not surveyed due to the security assessment by Integrity Watch. The 184 schools dropped from the sample and the 10 schools not surveyed after selection help gauge the impact security dynamics at the local level have in impeding progress in education.

### 2. Limitations

Notwithstanding the best efforts of the research team, some limitations remain. The findings in this survey are based on pre-selected schools in 10 provinces chosen to represent the whole country, which include 42 CERP-funded schools in five provinces, 77 Norwegian-funded schools in Faryab, and 157 USAID-funded schools in eight provinces. In Faryab, while the field survey covered a large sample, it could not cover areas with a fragile security situation. Hence, primary data for these districts remains limited. The research team sought to mitigate this limitation by gathering information for those regions through the provincial education directorate as well as from members of local communities from the surrounding areas. While the questionnaire was designed to encapsulate a range of key issues on which sustainability of schools is to be gauged, it does not provide us with specific information such as building standards followed or per-unit cost of construction. The survey team took great care to ensure accuracy in collecting data on school enrollment by age and gender. However, given the absence of documents such as birth certificates and attendance of girls in schools designated on paper for boys and vice versa owing to non-availability of any other schools nearby, figures represented here are to be read as proxy indicators. To the extent possible, this limitation was sought to be overcome by triangulating data generated.

## **Analysis of the Policy Issue**

The in-depth analysis of the policy issue is written here. This will likely be the largest section, setting the basis for the policy options you recommend. If the analysis of the policy issue is confusing or poorly structured, the policy options might not come across as justified or ‘rooted in reality’. It is important to convince the reader that the issue warrants attention and that the status quo is insufficient.

It is helpful to break this section up into themes and sub-themes that connect to one another. Establishing a compelling analysis will require you to draw on multiple sources, effectively present original research, and strategically package each theme.

### **Drawing on existing literature:**

When making claims about an issue, drawing on peer-reviewed literature adds credibility to your argument. If, for example, there is limited research on meritocratic police appointments in Afghanistan, one can draw on research on the importance of transparent and accountable police institutions for post-conflict reconstruction in general. It is important to ask yourself what implicit assumptions you are making about the policy issue, its significance or in some cases, even its existence. Drawing on findings from systematic literature reviews is often the gold standard when it comes to policy issues that cannot enjoy the predictive power of randomized control trials or establish a counterfactual. When this is not available, drawing on multiple sources with similarities to Afghanistan's location or status as a post-conflict society can be useful for strengthening claims.

### **Packaging themes:**

It is important to frame themes that are distinct yet interrelated. While you will likely come up with themes during the analysis portion of your research study, it is important that you package them with your target audience in mind. Consider adopting the terminology that your target audience uses to avoid confusion. It can also be helpful to have a similar structure for each subsequent theme. You can, for example, introduce the theme, identify the link between the theme and the policy issue, share key findings, and then discuss the implications of your findings on those identified linkages. This process can be repeated for each theme.

### **Presenting original research:**

It might be tempting to present primary research the same way that it was conducted (i.e., in chronological order or by region), but this is likely not the most strategic approach. Instead, organize findings by theme as they relate to the policy issue. The primary purpose of including original research in a policy paper is to fill an evidence gap and to justify the policy options one puts forward.

Make use of headings, lists, bolded text, and italics to guide the reader along. Incorporate tables, figures, or maps to give a visual of your data and make sure to draw the reader's attention to them at the appropriate point in the text. Do not include a figure, table, or map if you do not plan to refer to it. Consider using boxes, where you can do a 'deep-dive' on a linked but distinct issue, on a case study, or on an interview with an individual to get a more qualitative account of the phenomena involved.

See the shortened example below from IWA's [Education Compromised? A Survey of Schools in 10 Provinces of Afghanistan](#) (Sharma and Afzali 2018). The box takes a closer look at the issue of 'ghost schools' that, while important, is kept separate from the main text and included in its own text box. This maintains the flow of the overall text, while creating 'optional' cases that the reader can refer to for more information on a specific topic or case.

Figure 17: Example of a box

*Box 4: Limitations in tracking donor support to schools*

The issue of ghost schools, ghost teachers, and ghost students has been ringing bells for the last few years in Afghanistan. One of the fundamental reasons for a large number of schools existing on paper only is the lack of reliable data based on a proper inventory and tracking system of donor support to schools. USAID has been a key donor to the education sector and has spent hundreds of millions of dollars since 2002. Integrity Watch assessed 157 USAID-funded schools in eight provinces. All schools constructed through USAID funding are supposed to have the USAID emblem on the building. However, surveyors for this report were not able to locate the emblem anywhere in at least 63 schools in Herat, Balkh, Khost, Parwan, Baghlan, Faryab, and Kunduz.

## Policy Options

After the policy issue has been analyzed in the previous section and the ‘problem’ has been defined, one can present ‘solutions’ in the form of policy options. There is no one formula for creating policy options as every policy issue is unique. One can, for example, present a status-quo policy option, a reformative policy option with non-foundational changes, or a more ambitious transformative policy option. Another option might be to present options that coincide with the stances of different political parties on a policy issue.

Regardless of how one chooses to frame their set of policy options, a common evaluative framework should be used to assess each option against the researcher’s evidence base. One can conduct a SWOT analysis (strengths, weaknesses, opportunities, threats), assess the equity dimensions of different options, the costs/benefits, feasibility of the option, or the impact on stakeholders, institutions, and the environment, for example.

## Scenario-building

It can be useful to build worst-case, best-case, and most-likely scenarios (see scenario-building in the Tools section below) to come up with policy options in accordance with each scenario. This approach can strengthen the policy paper by incorporating contingencies that adapt as the context changes.

Scenarios are developed by assessing trends in key indicators, as well as among stakeholders. Once these trends are understood, it is possible to make a judgment on where "things are going" by weighing up (and against each other) indicators and stakeholder developments. The most useful scenarios in operational terms are those that lay out (a) the best-possible scenario that can be reached under current circumstances, (b) the middle-case scenario, which describes a “muddling through” outcome of the current situation, and (c) the worst-case scenario for which practitioners should be prepared. (“Early Warning - Early Response Handbook” 2013)

For operational purposes, it may be useful to provide an assessment of the likelihood of different scenarios (most likely, less likely, likely), as well as a time frame for their possible realization. The additional value of scenarios is that they are easily translated into overall objectives, thus “rooting” objectives in reality. As such, an optimal objective can be focused on realizing a best-

case scenario and contingency objectives can be focused on avoiding—and being prepared for—a worst-case scenario. (“Early Warning - Early Response Handbook” 2013)

The following are key elements of a scenario:

- Major events: Likely events that scenarios would be framed around (i.e. elections, negotiations, legislation timelines, significant cultural events).
- Time horizon: This determines the degree of predictability. Longer term scenarios are less predictable and demographic/technological/environmental trends take precedence over individual stakeholders or key events.
- Assumptions: High-probability, high impact factors that shape a scenario over a given timeframe based on evidence. This should cover political, economic, environmental, security, socio-cultural, and technological phenomena.
- Stakeholders/actors: Includes primary stakeholders with significant power to shape developments of interest and secondary stakeholders that tend to respond to other actors actions and who wield less influence in driving events. Local, national, or international stakeholders can be identified, along with interests and positive/negative linkages with other actors.
- Wildcards: Largely unpredictable elements in the issue’s context, but with the potential to significantly change the course of events and required contingencies (COVID-19).

### Policy Options and Recommendations

Policy options and recommendations are most robust with an understanding of the worst, best, and most likely scenarios. These scenarios provide entry points for action and justify what sets the recommended option apart from alternatives. In line with the three scenarios, it can be useful to consider three objectives:

- 1) Optimal objective: Direct efforts to realize the best-case scenario
- 2) Status quo objective: Avoid significant intervention, often to prevent the situation from getting worse
- 3) Contingency objective: Focus efforts on preventing the worst-case scenario from happening

Work through the following checklist to assess your policy options and the stakeholders they would implicate.

*Table 11: Stakeholder checklist*

<p><b>Who is my target audience?</b>  <i>Be as specific as possible, down to the name of politicians, departments of organizations, or media outlets</i></p>	
<p><b>What are their interests and positions?</b></p>	

<i>This should include official positions and third-party accounts of their intentions and motivations</i>	
<b>Do my policy options leverage synergies and mitigate conflicts among my target audience?</b>	
<b>Do my policy options address root causes or downstream effects?</b>	
<b>Do my policy options prevent the worst case, maintain the status quo, or work to realize the best case scenario?</b>	
<b>Is my recommendation specific and linked to an action that can be taken by particular stakeholders?</b>	
<b>Are there others reporting on this issue?</b> <i>Do their findings, or your hunch about their findings, compete with or complement your analysis?</i>	

**Recommendations and Conclusion**

This section should start with a brief synopsis of the policy analysis section and policy options discussed so far. Make sure to clearly identify your set of policy recommendations from the options assessed, where each policy recommendation addresses some component of the overall policy issue. You might want to outline these in a list with bolded subheadings, and wherever possible, connect recommendations with the appropriate decision-maker and set a time horizon (i.e. ahead of the upcoming election, prior to the withdrawal of foreign troops, etc.)

If necessary, you can conclude with closing remarks on the policy issue, reasserting its significance and making a call to action.

For example, consider the shortened excerpt from the recommendations in [Education Compromised? A Survey of Schools in 10 Provinces of Afghanistan](#) (Sharma and Afzali 2018) below. The section begins with a synopsis of the findings and introduces several recommendations, each with a bolded sub-heading. The author ties each recommendation to key decisionmakers and provides practical, tangible solutions that address each of the thematic problem areas that arose in the primary research.

**RECOMMENDATIONS**

Key findings of this report in relation to rates of enrollment, retention, and learning outcomes in schools have reverberations for the school education sector across the country given the broadly similar nature of

socio-cultural, economic, political and security challenges faced. Based on the key gaps in school education identified in this report the following set of recommendations is put forth for consideration:

**Improving physical infrastructure:** The MOE should explore ways of engaging communities in planning, construction, and maintenance of infrastructure by developing an institutional framework for effectively channelizing community involvement. Adequate resources should be allocated for maintenance of schools to ensure their sustainability. The community-based monitoring of schools is a model that could be adopted for this and other purposes to improve quality of education. Avenues for providing schools access to sources of (renewable) power should be explored in collaboration with local and/ or regional NGOs. In addition, establishment of a school maintenance fund for each school should be explored. This should include incentivized contribution from MOE but should also attract local resources such as community contributions. Such funds should be spent in a transparent manner including communities in decision-making and keeping the funds in a dedicated bank account.

**Developing standards and guidelines for school infrastructure:** The MOE, in consultation with Provincial Educational Directorates (PEDs) as well as civil society actors active in delivering educational services should come up with standardized construction, maintenance, and operation guidelines for schools that should among other things ensure access to sanitation, portable water, electricity, and boundary walls. The implementation of this could be overseen by a CBM structure with a line of reportage to the PEDs so as to ensure compliance with guidelines.

## Appendices and Annexes

Annexes and appendices are both additions placed at the end of the policy paper, but before the bibliography. Annexes are stand-alone documents while appendices will usually not be meaningful outside the context of the policy paper. Often, the author of an annex is different from the author of the policy paper, but the author of an appendix is often the author of the policy paper.

Both annexes and appendices supplement the essential components of your policy paper by providing more detailed information. This information, while important, would derail the focus of the paper or distort the format more than it would help if included within the paper. Include only essential information in the appendices and annexes and avoid ‘dumping’ loosely connected data tables, research notes, or figures. If you do not reference an appendix or annex in the main body of your policy paper, you likely do not need to include it.

Appendices and annexes are usually titled with letters corresponding to their order and are listed in the table of contents.

## Citation guide

IWA publications mostly use *The Chicago Manual of Style (CMOS)* when referencing content. CMOS has two formats, Author-Date and Notes-Bibliography (NB). Each have their unique styles and use-cases, but NB is mostly used in IWA publications. These guidelines for NB below draw from [OWL Purdue](#), please visit the website for in-depth guidelines on other referencing styles (“Chicago Manual of Style 17th Edition // Purdue Writing Lab” n.d.)

## Guidelines for *The Chicago Manual of Style* (17<sup>th</sup> edition), Notes-Bibliography system

- Margins should be no less than 1”;
- Standard font (Times New Roman or Courier), equal or more than 10-point size;
- No specific requirements for subheadings, but should be consistent;
- Superscript numbers (i.e.,<sup>1</sup>) refer to a note and must come at the end of a sentence’s punctuation. For example, note the superscript in the following sentence:
  - The report found no evidence to support the government’s claim.<sup>2</sup>
- When an article that you are referencing has one, two, or three authors, always list all their names in the bibliography and notes. When an article has four to ten authors, all authors should have their names written in the bibliography entry, but only the first author’s name and “et al.” should be used in the notes;
- Either footnotes or endnotes can be used. Footnotes (sources at the bottom of each page) are preferable but endnotes (all sources in a notes section at the end of the document just before the bibliography) can be used if footnotes take a large amount of space in the document;
- Capitalize titles of texts mentioned in the main body text, notes, and in the bibliography; and
- Book titles and journal names are italicized, while articles and chapter titles are enclosed in double quotation marks.

**Citations:** NB uses numbered footnotes when referencing content. There is a reference at the bottom of the page that refers to an entry in the bibliography. The following elements are used in a footnote and bibliographical entry:

Author name	Title	Journal	Volume Number	Issue Number	Date	Page Number	URL/DOI
----------------	-------	---------	------------------	-----------------	------	----------------	---------

### Footnote:

<sup>1</sup>Jane Doe, “Title of work”, *Journal* 2, no. 2 (January 1): 20-24, www.google scholar.com

### Bibliography:

Doe, Jane. “Title of work.” *Journal* 2, no. 2 (January 1): 20-24. www.google scholar.com.

The footnote and bibliography entries are identical, except that

- Each footnote begins with a number starting at 1, and can be in superscript in the footnote itself;
- Elements in the bibliography are separated by periods while elements in the footnote are separated by commas;

- In the bibliography, the first author's name is inverted (i.e. Doe, Jane) but the rest of the authors' names are not. In the footnote, names are not inverted (i.e. Jane Doe); and
- Page numbers refer only to relevant sub-sections in the original text in the footnote but refer to the article's first and last page in the bibliography.

Example 1 (see note below)

The report found that the program had no effects on attitude towards women.(Beath, Christia, and Enikolopov 2013)

The footnote is colour-coded to match the format above.

<sup>5</sup>Andrew Beath, Fotini Christia, and Ruben Enikolopov, "Empowering Women through Development Aid: Evidence from a Field Experiment in Afghanistan," *American Political Science Review* 107, no. 3 (August 2013): 540–57, <https://doi.org/10.1017/S0003055413000270>.

The bibliographical entry for this footnote follows below. Note the inverted last names and the use of periods to separate elements instead of commas:

Beath, Andrew, Fotini Christia, and Ruben Enikolopov. "Empowering Women through Development Aid: Evidence from a Field Experiment in Afghanistan." *American Political Science Review* 107, no. 3 (August 2013): 540–57. <https://doi.org/10.1017/S0003055413000270>.

In the following example, the subsequent citation is shortened as it has been noted already. The use of *Ibid.* is discouraged in the 17<sup>th</sup> edition of Chicago NB and instead a shortened citation with the name of the author, and if possible the page number(s) is encouraged.

Example 2 (see note below):

The study failed to reject the null hypothesis.<sup>1</sup>

Footnote:

<sup>1</sup> Beath, Christia, and Enikolopov, 3.

Your bibliography should be titled Bibliography, with two blank lines below the title and one blank line between each source. Sources should have a 1 inch hanging indent and be sorted alphabetically letter by letter. In the following example, note how when two articles start with the same letter, they are sorted by the second letter.

### Example 3

#### Bibliography

---

<sup>1</sup> Beath, Christia, and Enikolopov, 3.

- Afzali, Sayed Ikram, and Mohammad Naser Timory. 2017. "Policy Brief Review of Government's Anti-Corruption.Pdf." Integrity Watch Afghanistan. <https://iwaweb.org/wp-content/uploads/2014/12/Policy-brief-Review-of-Government%E2%80%99s-Anti-Corruption.pdf>.
- Allan, Emily, Abdul Najm, Helen Fernandes, and Becca Allchin. 2018. "Participatory Lived Experience Research: Barriers and Enablers for Social Inclusion for People with Psychosocial Disability, in Afghanistan." *Intervention* 16 (3): 222–30. [https://doi.org/10.4103/INTV.INTV\\_39\\_18](https://doi.org/10.4103/INTV.INTV_39_18).
- Barakat, Sultan, and Gareth Wardell. 2002. "Exploited by Whom? An Alternative Perspective on Humanitarian Assistance to Afghan Women." *Third World Quarterly* 23 (5): 909–30. <https://doi.org/10.1080/0143659022000028585>.
- Beath, Andrew, Fotini Christia, and Ruben Enikolopov. 2013. "Empowering Women through Development Aid: Evidence from a Field Experiment in Afghanistan." *American Political Science Review* 107 (3): 540–57. <https://doi.org/10.1017/S0003055413000270>.

For more details on referencing, please consult the [Owl Purdue resource](#) ("Chicago Manual of Style 17th Edition // Purdue Writing Lab" n.d.).

## Innovative Tools

This section will incorporate innovative tools into the report-writing 'workflow' using Microsoft Word and free tools including: [Zotero](#) for reference management; [NVivo](#) for qualitative data coding, storage, and analysis; [Small SEO Tools](#) to check for plagiarism; [grammarly](#) to check for spelling, grammar, punctuation, word choice, tone, and style; [Pixabay/Pexels](#) for free stock images; and [Canva](#) for graphics and report templates.

**Reference Management with Zotero:** Manually compiling, storing, organizing, and citing references can take a long time. Longer reports with many references can become complex to manage and are prone to human error. Reference management software like [EndNote](#) (paid), [Mendeley](#) (free), and [Zotero](#) (free) simplify that process. These tools scrape meta-information (i.e. author name, title, journal, date, etc.), download article PDFs, keep your references organized, and create a formatted bibliographical entry for you.

Zotero is free and open-source, with paid options for more storage. It features a web browser extension that, with one-click, scrapes meta-information and the article PDF to write a bibliographical entry for you. Its integration with Microsoft Word automatically writes in-text citations for articles you enter in the search bar and in the reference style of your choice. If, for example, you need to change the citation style for the entire report after you have finished a first draft, you can do so for all citations with one-click.

Zotero allows you to work collaboratively with others, whereby your team can contribute to the same shared dataset of articles or even create a webpage where your dataset can be made public

or private. Exceeding the limit on free storage (300 MB) will require you to pay a small fee to continue adding articles on your designated webpage.

**Qualitative Research with NVivo:** Coding qualitative data by hand or with software that is not designed for the task can take a long time and make it difficult to stay organized, search through codes, replicate one's process, and form thematic linkages. Tools like NVivo (paid) and Atlas.ti (paid) simplify this process. Most qualitative data analysis software is proprietary and free options are limited. [QDA Miner Lite](#) is an example, although with functionality limited to code and retrieve operations.

NVivo allows the user to import word documents, PDFs, videos, pictures, audio, spreadsheets, and other file types. Integration with reference management software like Zotero, EndNote, and Mendeley makes importing literature easy and quick. Once imported, files are coded by highlighting segments of text and assigning them to thematic codes, cases, and sets. Codes are thematic (i.e. "Maternal Health" and "Mental Health" would be two distinct codes). Cases refer to particular persons, places or distinct items of interest.

In IWA's report on [Education Compromised? A Survey of Schools in 10 Provinces of Afghanistan](#) (Sharma and Afzali 2018) each school would be a case, and all findings on the variables of interest (operational status of schools, quality of infrastructure, access to sanitation, etc...) over the three-year timespan would be included for each school, along with relevant demographic information. These cases then become the 'units of analysis' in a study.

Sets refer to groupings of files and/or codes, useful when splitting work among a team or gathering information in multiple waves.

Once all files are coded, the data can be explored in several ways. One can:

- Compare how people talk about a particular topic based on demographic or other differences (i.e. comparing how people of different genders or different positions within a system responds to a set of questions)
- Query the codes to identify linkages (i.e., searching for text that is coded both as "Mental Health" and "Education");
- Automatically assess sentiment based off word-choice (i.e., Very negative, moderately negative, moderately positive, very positive sentiment among interview responses to a new city policy);
- Create relationships between nodes (i.e., The node "Insecurity" could be connected by the special node "Increases" to the node "Migration". It allows the relationship between two nodes to be captured in a separate node.); or
- Create visualizations to answer complex questions (i.e., comparison diagrams, word clouds, word frequency charts).

NVivo has a collaboration version where teammates can work together on a project that is stored online. A team can also work together offline without the collaboration version by creating a master version with all files and an initial node structure, with a 'set' of files, cases, or nodes for

each teammate. Everyone can work on their own designated set and when finished, each teammate's project files can be merged to create a final document. Note that if teammates are working on different operating systems (Mac and Windows), they will need to take the extra step of converting the file to one of these versions before merging. They should also be mindful of slight differences in coding due to how Mac and Windows handle blank spaces and avoid using features that the [Windows version has that the Mac version does not](#) (i.e. compound and group queries, see-also links, dynamic sets, framework matrices, sentiment coding, relationship coding, etc.,).

NVivo also has a separate [transcription service](#) in 28 languages that automatically transcribes audio recordings of interviews or focus group discussions but does not include Pashto and Dari.

For further reading, the following sources are recommended:

Bazeley, P. (2018). Using qualitative data analysis software to assist data analyses. In P. Liamputtong (Ed.), *Handbook of research methods in health social sciences*. Singapore: Springer.

Bazeley, P. (in press). Using NVivo for mixed methods research. In A. Onwuegbuzie & R. B. Johnson (Eds.), *Reviewer's guide for mixed methods data analysis*. New York, NY: Routledge.

### **Quantitative Research with R and RStudio**

R is a free and open-source [programming language](#) widely used for quantitative analysis involving both descriptive and interpretive statistics. The language is commonly used in conjunction with [RStudio](#), which provides a graphical interface and manages many technical add-ons behind the scenes for the user.

You can import and view datasets from a range of filetypes (i.e. .csv, .txt, .xls, .xlsx, SPSS, Stata, and SAS files), perform simple and complex statistical operations (i.e. measures of central tendency, standard deviation, linear and logistical regression, ANOVA, etc.,) and visualize your data with customizable graphics (box plots, histograms, QQ plots, etc.,).

You can take advantage of hundreds of free community-sourced add-ons called 'packages' that provide more complex statistical methods, virtual trainings, and visualizations. Perhaps most compelling is the emergence of big data and the ability to use R for deep learning. The applicability to policy issues is nascent but an area of future development.

There are a number of online tutorials available on how to use R for quantitative analysis, please see [swirl](#), [R Programming A-Z](#), and [Quantitative Social Science: An Introduction](#) by Kosuke Imai as highly recommended options.

### **Proofreading and plagiarism with Small SEO Tools and Grammarly**

The free version of [Grammarly](#) provides more useful suggestions on grammar, punctuation, conciseness, and spelling than the default editor on Microsoft Word or other platforms. The free browser extension provides suggestions when writing emails on Gmail and Outlook, sending messages or posting on social media, and working on project management platforms like Slack, Asana, or Salesforce.

[Small SEO tools](#) also provides numerous functions including a spelling, grammar, and plagiarism checker. The plagiarism checker rapidly cross-references text that you write with billions of web-pages to identify any matches, leaving the writer and editor confident that their content is original. There is also a paraphrasing and article rewriter tool that automatically rephrases content, saving time for the report-writer when citing longer blocks of text or summarizing key information.

### **Graphic Design with Canva, Pixabay, and Pexels**

[Pixabay](#) and [Pexels](#) provide thousands of free high-resolution stock photos and videos for commercial and non-commercial use, without requiring attribution and with minimal restrictions.

[Canva](#) provides templates, unique fonts, digital icons, and a simple platform to create fully customizable reports with a professional look and feel. Apart from some in-app purchases of premium features, Canva is free to use for non-profit organizations. The webapp has teamwork functionality, allows for customized file exports and can even be used to create unique, interactive presentation slide decks.

# References

## Program and Impact Evaluation References and links

- Austin, Alex. 2003. *Peace and Conflict Impact Assessment Critical Views on Theory and Practice. Broadening the Range Sharing the of Designs and Benefits Of Trade For Impact Evaluations.* 38, Department for International Development, 2012, <https://assets.publishing.service.gov.uk/media/57a08a6740f0b6497400059e/DFIDWorkingPaper38.pdf>
- Bush, Kenneth, and Colleen Duggan. 2015. *Evaluation in the Extreme: Research, Impact and Politics in Violently Divided Societies* <https://www.deslibris.ca/ID/10063069>.
- Chapter 2: *Conflict Analysis*. Conflict Sensitivity Resource Pack, 2015, [https://conflictsensitivity.org/resource\\_pack/chapter-2-conflict-analysis/](https://conflictsensitivity.org/resource_pack/chapter-2-conflict-analysis/)
- Church, Madeline (2019). *Doing Things Differently: Rethinking Monitoring and Evaluation to Understand Change* <https://www.saferworld.org.uk/resources/publications/1027-doing-things-differently-rethinking-monitoring-and-evaluation-to-understand-change>
- Conflict Sensitivity: Experiences from Local and Community Development Practice in Myanmar. 2017. UNDP [https://www.international-alert.org/sites/default/files/UNDP\\_ConflictSensitivityPracticeGeneral\\_EN\\_2017.pdf](https://www.international-alert.org/sites/default/files/UNDP_ConflictSensitivityPracticeGeneral_EN_2017.pdf)
- Conflict Sensitivity: Experiences from UNDP's Local & Community Development Practice. 2017. UNDP [https://www.internationalalert.org/sites/default/files/UNDP\\_ConflictSensitivityPracticeUNDP\\_EN\\_2017.pdf](https://www.internationalalert.org/sites/default/files/UNDP_ConflictSensitivityPracticeUNDP_EN_2017.pdf)
- Conflict Sensitivity: Indicators for Local and Community Development Programming in Myanmar. 2017. UNDP [https://www.internationalalert.org/sites/default/files/UNDP\\_ConflictSensitivityIndicators\\_EN\\_2017.pdf](https://www.internationalalert.org/sites/default/files/UNDP_ConflictSensitivityIndicators_EN_2017.pdf)
- Country Indicators for Foreign Policy Reports on Fragile and Conflict Affected States. <https://carleton.ca/cifp/failed-fragile-states/reports-and-briefs/>
- Digging Deeper With The People: A Survey of How Corruption in the Justice System in Afghanistan Affects Citizens' trust. Integrity Watch Afghanistan. 2020. IWAWEB <https://iwaweb.org/blog/publication/extractive-sector/digging-deeper-with-the-people-a-survey-of-how-corruption-in-the-justice-system-in-afghanistan-affects-citizens-trust/>
- Doing Research in Conflict Settings: Gender Mainstreaming and Ethics. 2019. *Safe World* <https://www.saferworld.org.uk/resources/publications/1209-doing-research-in-conflict-settings-gender-mainstreaming-and-ethics>
- Early Warning - Early Response Handbook*. Peacebuilding Centre. 2013. *Peacebuilding Center*

- [http://peacebuildingcentre.com/psc\\_documents/ewhandbook\\_eng.pdf](http://peacebuildingcentre.com/psc_documents/ewhandbook_eng.pdf)  
Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners (2000). *World Bank Group* <http://www.myilibrary.com?id=9072>.
- Handbook on Monitoring and Evaluating for Results. 2004. *UNDP Evaluation Office* <http://web.undp.org/evaluation/documents/HandBook/ME-Handbook.pdf>
- How to Guide to Conflict Sensitivity. 2012. *Conflict Sensitivity Consortium* [https://conflictsensitivity.org/wp-content/uploads/2015/04/6602\\_HowToGuide\\_CSF\\_WEB\\_3.pdf](https://conflictsensitivity.org/wp-content/uploads/2015/04/6602_HowToGuide_CSF_WEB_3.pdf).
- Impact Evaluation in UN Agency Evaluation Systems: Guidance on Selection, Planning and Management. (2013). *UNEG Impact Evaluation Task Force, Aug.*
- Joyce, Craig M., et al (2018). *Measuring Impact by Design: A Guide to Methods for Impact Measurement* [http://epe.lac-bac.gc.ca/100/201/301/weekly\\_acquisitions\\_list-ef/2019/19-18/publications.gc.ca/collections/collection\\_2019/bcp-pco/CP22-174-2019-eng.pdf](http://epe.lac-bac.gc.ca/100/201/301/weekly_acquisitions_list-ef/2019/19-18/publications.gc.ca/collections/collection_2019/bcp-pco/CP22-174-2019-eng.pdf)
- Khandker, Shahidur R., et al. 2010. Handbook on Impact Evaluation: Quantitative Methods and Practices. *World Bank* <http://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=476187>
- MSH (Management Sciences for Health), and UNICEF (United Nations Children’s Fund). “Quality Guide: Stakeholder Analysis.” In *Guide to Managing for Qualité* <http://bsstudents.uce.ac.uk/sdrive/Martin%20Beaver/Week%20202/Quality%20Guide%20Stakeholder%20Analysis.htm>
- Morra-Imas, Linda G., and Ray C. Rist. 2009. *The Road to Results: Designing and Conducting Effective Development Evaluations*. *World Bank*.
- Ofir, Zenda, et al. 2016. *Research Quality Plus A Holistic Approach to Evaluating Research*. International Development Research Centre <https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/56528/IDL-56528.pdf?sequence=2&isAllowed=y>
- Tarrow S. 2009. Bridging the Quantitative-Qualitative Divide. In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, Brady HE and Collier D (eds). Lanham, MD: Rowman and Littlefield.
- UNDP Evaluation Guidelines. 2019. Independent Evaluation Office of UNDP. <http://web.undp.org/evaluation/guideline/>
- Wholey, Joseph S., et al (2010). *Handbook of Practical Program Evaluation*. <http://www.vlebooks.com/vleweb/product/openreader?id=none&isbn=9780787973889>

Need Analyses References and links

Watkins, Ryan, et al. 2012. *A Guide to Assessing Needs: Essential Tools for Collecting*

*Information, Making Decisions, and Achieving Development Results.* World Bank.  
<http://documents1.worldbank.org/curated/ar/644051468148177268/pdf/663920PUB0EPI00essing09780821388686.pdf>

#### Peace and Conflict Impact Assessment References and links

Austin, Alex. 2003. *Peace and Conflict Impact Assessment Critical Views on Theory and Practice*  
<https://berghof-foundation.org/library/peace-and-conflict-impact-assessment-critical-views-on-theory-and-practice>

Conflict Sensitivity: Experiences from Local and Community Development Practice in Myanmar. 2017. *UNDP*

[https://www.international-alert.org/sites/default/files/UNDP\\_ConflictSensitivityPracticeGeneral\\_EN\\_2017.pdf](https://www.international-alert.org/sites/default/files/UNDP_ConflictSensitivityPracticeGeneral_EN_2017.pdf)

Conflict Sensitivity: Experiences from UNDP's Local & Community Development Practice. 2017. *UNDP*

[https://www.internationalalert.org/sites/default/files/UNDP\\_ConflictSensitivityPracticeUNDP\\_EN\\_2017.pdf](https://www.internationalalert.org/sites/default/files/UNDP_ConflictSensitivityPracticeUNDP_EN_2017.pdf)

Conflict Sensitivity: Indicators for Local and Community Development Programming in Myanmar. 2017. *UNDP*

[https://www.internationalalert.org/sites/default/files/UNDP\\_ConflictSensitivityIndicators\\_EN\\_2017.pdf](https://www.internationalalert.org/sites/default/files/UNDP_ConflictSensitivityIndicators_EN_2017.pdf)

Chigas, Diana, Madeline Church, and Vanessa Corlazzoli. 2014. "Evaluating Impacts of Peacebuilding Interventions: Approaches and Methods, Challenges and Considerations." *A Conflict, Crime, and Violence Results Initiative (CCVRI) product*. London, UK: DFID, <https://www.cdacollaborative.org/publication/evaluating-impacts-of-peacebuilding-interventions-approaches-and-methods-challenges-and-considerations/>

Menkhaus, Ken. 2004. *Impact assessment in post-conflict peacebuilding*

<https://www.interpeace.org/resource/impact-assessment-in-post-conflict-peacebuilding/>

#### Sampling References and links

Acharya, Anita & Anupam Prakash, Pikee Saxena, Aruna Nigam. 2013. "Sampling: Why and How of It?" *Indian Journal of Medical Specialties* 4, no. 2: 330-333.

Adib, Ezatullah & Andrew McDevitt. 2018. "National Corruption Survey 2018: Afghans' Perceptions and Experiences of Corruption." *IWA* [https://iwaweb.org/wp-content/uploads/2014/12/NCS\\_2018\\_English\\_WEB.pdf](https://iwaweb.org/wp-content/uploads/2014/12/NCS_2018_English_WEB.pdf)

Bacchetti, Peter & Leslie E. Wolf, Mark R. Segal, and Charles E. McCulloch. 2004.

"Ethics and Sample Size." *American Journal of Epidemiology* 61, no. 2: 105-110.

- Berdahl, Loleen & Kieth Archer. 2015. *Explorations: Conducting Empirical Research in Canadian Political Science*. Don Mills: Oxford University Press.
- Boyd, Paul. 2006. "By the Numbers: A Sample Size Table." *Quirk's Media*.  
<https://www.quirks.com/articles/by-the-numbers-a-sample-size-table>
- Cohen, Nissim & Tamar Arieli. 2011. "Field Research in Conflict Environments: Methodological Challenges and Snowball Sampling." *Journal of Peace Research* 48, no. 4: 423–435.
- Cronin-Furman, & milli Kate & Lake. 2018. "Ethics Abroad: Fieldwork in Fragile and Violent Contexts." *Political Science and Politics* 51, no. 3: 607-614.
- Dattalo, Patrick. 2010. "Ethical Dilemmas in Sampling." *Journal of Social Work Values and Ethics* 7, no. 1.
- Daniel, Johnnie 2012. *Sampling Essentials: Practical Guidelines for Making Sampling Choices* 12  
margin of error?" *Scientific America*.
- Haan, Michael & Jenny Godley. 2017. *An Introduction to Statistics for Canadian Social Scientists*, Third Edition. New York: Oxford University Press.
- Halperin, Sandra & Oliver Heath. 2017. *Political Research: Research and Practical Skills*. New York: Oxford University Press.
- Ibrahimi, Yaqub. 2020. "Political Settlement and Post-Conflict Order in Afghanistan: People's Views." *Afghan Institute for Strategic Studies*.
- Lauren, Thomas. 2020. "An Introduction to Simple Random Sampling." *Scriber*  
<https://www.scribbr.com/methodology/simple-random-sampling/#:~:text=Simple%20random%20sampling%20is%20a,possible%20of%20this%20Random%20subset>
- Sample Size Calculator. 2020. *Qualtrics*  
<https://www.qualtrics.com/blog/calculating-sample-size/>
- Young, An Gie & Sean Pearce. 2013. "A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis." *Tutorials in Quantitative Methods for Psychology* 9, no. 2: 79-94.
- Schreier, Margrit. 2014. "Qualitative Content Analysis." In Uwe Filk ed. *The Sage Handbook of Qualitative Data Analysis*. London: Sage.

#### Surveys References and links

- Akseer, Tabasum et al. 2019. "A Survey of the Afghan People." *Asia Foundation*.
- ASA 1995. *Best Practices for Survey Research*  
<https://www.aapor.org/Standards-Ethics/Best-Practices.aspx#best2>
- Berdahl, Loleen & Kieth Archer. 2015. *Explorations: Conducting Empirical Research in Canadian Political Science*. Don Mills: Oxford University Press.

- Canadian Election Study. 2015: Codebook and Questionnaire.  
[https://ceseec.sites.olt.ubc.ca/files/2017/04/CES2015\\_Combined\\_Data\\_Codebook.pdf](https://ceseec.sites.olt.ubc.ca/files/2017/04/CES2015_Combined_Data_Codebook.pdf)
- Check, J, & R.K. Schutt. 2012. "Survey research." In: J. Check, R. K. Schutt., editors. *Research methods in education* (159–185). Thousand Oaks, CA: Sage Publications.
- CIFP, *Democracy and Governance*. Carleton University.
- Cohen, Nissim & Tamar Arieli. 2011. "Field Research in Conflict Environments: Methodological Challenges and Snowball Sampling." *Journal of Peace Research* 48, no. 4: pp. 423–435.
- Cronin-Furman, & milli Kate & Lake. 2018. "Ethics Abroad: Fieldwork in Fragile and Violent Contexts." *Political Science and Politics* 51, no. 3: 607-614.
- De Groeve, Tom & Daniel Frederik, Mandrella Sara, Pardo Serrano, Luca Vernaccini. 2015. "Global Conflict Risk Index (GCRI) – Handbook for data and statistical analysis." *JRC Science and Policy Report*.
- Glasow, Priscilla. 2005. "Fundamentals of Survey Research Methodology." MITRE  
[https://www.mitre.org/sites/default/files/pdf/05\\_0638.pdf](https://www.mitre.org/sites/default/files/pdf/05_0638.pdf)
- Great Brook. <https://greatbrook.com/survey-question-types-when-to-use-each/>
- Haan, Michael & Jenny Godley. 2017. *An Introduction to Statistics for Canadian Social Scientists*, Third Edition. New York: Oxford University Press.
- Halperin, Sandra & Oliver Heath. 2017. *Political Research: Research and Practical Skills* New York: Oxford University Press.
- ICC/ESOMAR (2016). "ICC/ESOMAR International Code on Market, Opinion, Social Research, and Data Analytics." [https://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ICCESOMAR\\_Code\\_English\\_.pdf](https://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ICCESOMAR_Code_English_.pdf)
- Ibrahimi, Yaqub. 2020. "Political Settlement and Post-Conflict Order in Afghanistan: People's Views." *Afghan Institute for Strategic Studies*.
- Pinsonneault A. & K.L Kraemer. 1993. "Survey research methodology in management information systems: An assessment." *Journal of Management Information Systems* 10: 75-105.
- Singleton R.A. & B.C. Straits. 2009. *Approaches to social research*. New York: Oxford University Press.
- WAPOR. 2011. *WAPOR Code of Ethics*. <https://wapor.org/about-wapor/code-of-ethics/>

#### Focus Groups References and links

- Barbolet et al. 2005. *Vancouver Food System Assessment*.  
<https://foodsecurecanada.org/sites/foodsecurecanada.org/files/vanfoodassessrpt.pdf>
- Berdahl, Loleen & Kieth Archer 2015. *Explorations: Conducting Empirical Research in Canadian Political Science*. Don Mills: Oxford University Press.

- Carbert, Louise. 2003. "Above the Fray: Rural Women Leaders on Regional Development and Electoral Democracy in Atlantic Canada." *Canadian Journal of Political Science* 36, no. 1: 159-183.
- Carey, Martha Ann & Jo-Ellen Asbury. 2012. *Focus Group Research*. Walnut Creek, CA: Left Coast Press.
- Charmaz, K. 2001. Grounded theory. In R. M. Emerson (Ed.), *Contemporary field research: Perspectives and formulations*, 2nd ed (335–52). Prospect Heights, IL: Waveland Press
- Corbin, J., & A. Strauss. 2008. *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Thousand Oaks, CA: Sage
- Cronin-Furman, & milli Kate & Lake. 2018. "Ethics Abroad: Fieldwork in Fragile and Violent Contexts." *Political Science and Politics* 51, no. 3: 607-614.
- Drisco, J. & T Maschi. 2015. *Textual Analysis*. New York: Oxford University Press.
- Elliott, R. & L. Timulak. 2005. "Descriptive and Interpretive Approaches to Qualitative Research). In J, Miles & P. Gilbert eds (147-160). *A Handbook of Research Methods for Clinical and Health Psychology*. New York: Oxford University Press.
- Grbich, C. 2007. *Qualitative data analysis: An introduction*. Thousand Oaks, CA: Sage.
- Gupta, Dipak (2001). *Analyzing Public Policy: Concepts, Tools, and Techniques*. Thousand Oaks: CQ Press.
- Halperin, Sandra & Oliver Heath. 2017. *Political Research: Research and Practical Skills*. New York: Oxford University Press.
- IBM SPSS Statistics. <https://www.ibm.com/products/spss-statistics>
- Karimi, Mohammad Daud. 2016. "Afghan Community Consultation: Focus Group Discussion of Mental Health." *Link: Health and Community* <http://www.linkhc.org.au/wp-content/uploads/2017/05/Afghan-Community-Focus-Group-Discussion-on-Mental-Health-Report-FINAL.pdf>
- Knott, Eleanor. 2019. "Beyond the Field: Ethics after Fieldwork in Politically Dynamic Context." *Perspectives on Politics* 17, no. 1: 140-153.
- Lewins, A., & Silver, C. 2007. *Using software in qualitative research: A step-by-step guide*. London: Sage.
- Lezaun, Javier. 2007. "A Market of Opinions: The Political Epistemology of Focus Groups." *The Sociological Review* 55, no. 2: 130-151.
- Luntz, Frank. 1994. "Focus Group Research in American Politics." *The Polling Report*. <https://www.pollingreport.com/focus.htm>
- Morgan, David & Kim Hoffman. 2018. "Focus Group." In Uwe Flick ed. *The Sage Handbook of Qualitative Data Analysis*. London: Sage.
- Onwuegbuzie, Anthony et al. 2009. "A Qualitative Framework for Collecting and Analyzing Data in Focus Group Research." *International Journal of Qualitative Research* 8, no. 3: 1-21.

Richards, L., & J. M. Morse. 2007. *Readme first for a user's guide to qualitative methods* (2nd ed.). Thousand Oaks, CA: Sage

Saldana, Johnny. 2013. *The Coding Manual for Qualitative Researchers*. London: Sage.

Sim, Julius & Jackie Waterfield (2019). "Focus Group Methodology: Some Ethical Challenges." *Quality and Quantity* 53: 3003-3022.

Stewart, David & Prem N. Shamdasani, Dennis W. Rook. 2007. *Focus Group: Theory and Practice* London: Sage.

Text Analysis with NVivo 11 for Windows.

[https://www.youtube.com/watch?v=SgY0MNdKVwM&ab\\_channel=NVivobyQSR](https://www.youtube.com/watch?v=SgY0MNdKVwM&ab_channel=NVivobyQSR)

#### Policy Briefs and Policy Papers References and links

Afzali, Sayed Ikram, and Mohammad Naser Timory. 2017. "Policy Brief Review of Government's Anti-Corruption.Pdf." Integrity Watch Afghanistan. <https://iwaweb.org/wp-content/uploads/2014/12/Policy-brief-Review-of-Government%E2%80%99s-Anti-Corruption.pdf>.

Allan, Emily, Abdul Najm, Helen Fernandes, and Becca Allchin. 2018. "Participatory Lived Experience Research: Barriers and Enablers for Social Inclusion for People with Psychosocial Disability, in Afghanistan." *Intervention* 16 (3): 222–30. [https://doi.org/10.4103/INTV.INTV\\_39\\_18](https://doi.org/10.4103/INTV.INTV_39_18).

Barakat, Sultan, and Gareth Wardell. 2002. "Exploited by Whom? An Alternative Perspective on Humanitarian Assistance to Afghan Women." *Third World Quarterly* 23 (5): 909–30. <https://doi.org/10.1080/0143659022000028585>.

Beath, Andrew, Fotini Christia, and Ruben Enikolopov. 2013. "Empowering Women through Development Aid: Evidence from a Field Experiment in Afghanistan." *American Political Science Review* 107 (3): 540–57. <https://doi.org/10.1017/S0003055413000270>.

Boucher, Paloma, Sarah DelVillano, Liu Shi, and Maseh Hadaf. 2020. "2020 Afghanistan Conflict Diagnostic." iAffairs Canada. <https://carleton.ca/cifp/wp-content/uploads/2020-Afghanistan-Conflict-Diagnostic-.pdf>.

"Chicago Manual of Style 17th Edition // Purdue Writing Lab." n.d. Purdue Writing Lab. Accessed February 5, 2021.

[https://owl.purdue.edu/owl/research\\_and\\_citation/chicago\\_manual\\_17th\\_edition/cmoss\\_formatting\\_and\\_style\\_guide/chicago\\_manual\\_of\\_style\\_17th\\_edition.html](https://owl.purdue.edu/owl/research_and_citation/chicago_manual_17th_edition/cmoss_formatting_and_style_guide/chicago_manual_of_style_17th_edition.html).

"Collaborate & Create Amazing Graphic Design for Free." n.d. Accessed February 5, 2021. <https://www.canva.com/>.

"Download the RStudio IDE." n.d. Accessed February 5, 2021. <https://rstudio.com/products/rstudio/download/>.

"Early Warning - Early Response Handbook." 2013. Peacebuilding Centre. [http://peacebuildingcentre.com/psc\\_documents/ewhandbook\\_eng.pdf](http://peacebuildingcentre.com/psc_documents/ewhandbook_eng.pdf).

- “EndNote | Clarivate Analytics.” n.d. EndNote. Accessed February 5, 2021. <https://endnote.com/>.
- FAO. n.d. “M&E Technical Advisory Note: Overview of Methods for Baseline Assessments.” Accessed October 19, 2020. <http://www.fao.org/3/a-az931e.pdf>.
- “Free Qualitative Data Analysis Software | QDA Miner Lite.” n.d. *Provalis Research* (blog). Accessed February 5, 2021. <https://provalisresearch.com/products/qualitative-data-analysis-software/freeware/>.
- Grant, Maria J., and Andrew Booth. 2009. “A Typology of Reviews: An Analysis of 14 Review Types and Associated Methodologies.” *Health Information & Libraries Journal* 26 (2): 91–108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>.
- Johnston, Ron, and Paul Plummer. 2005. “What Is Policy-Oriented Research?” *Environment and Planning A* 37 (February): 1521–26. <https://doi.org/10.1068/a3845>.
- Kim, Christine, Khwaja Mir Ahad Saeed, Ahmad Shah Salehi, and Wu Zeng. 2016. “An Equity Analysis of Utilization of Health Services in Afghanistan Using a National Household Survey.” *BMC Public Health* 16 (1): 1226. <https://doi.org/10.1186/s12889-016-3894-z>.
- Leacock, Dr Vivian. 2013. “Public Procurement Reform - A Rapid Evidence Review.” Scottish Government. <https://www.webarchive.org.uk/wayback/archive/3000/https://www.gov.scot/resource/0043/00430602.pdf>.
- Mackenzie, Jean. 2016. “National Corruption Survey 2016: Afghans’ Perceptions and Experiences of Corruption.” Integrity Watch Afghanistan. [https://iwaweb.org/wp-content/uploads/2016/12/NCS\\_\\_English\\_\\_for-web.pdf](https://iwaweb.org/wp-content/uploads/2016/12/NCS__English__for-web.pdf).
- Mclvor, Mitchell. 2018. “Writing Policy Briefs & Reports Overview, Tips, & Resources.” [https://www.utm.utoronto.ca/asc/sites/files/asc/public/shared/pdf/wdi/sample\\_course\\_materials/soc/SOC\\_PolicyWritingGuide.pdf](https://www.utm.utoronto.ca/asc/sites/files/asc/public/shared/pdf/wdi/sample_course_materials/soc/SOC_PolicyWritingGuide.pdf).
- “Mendeley Reference Manager | Mendeley.” n.d. Accessed February 5, 2021. <https://www.mendeley.com/reference-management/reference-manager/>.
- Miller, Delbert, and Neil Salkind. 2002. *Handbook of Research Design & Social Measurement*. 2455 Teller Road, Thousand Oaks California 91320 United States of America: SAGE Publications, Inc. <https://doi.org/10.4135/9781412984386>.
- “Pexels.” n.d. Accessed February 5, 2021. <https://www.pexels.com/>.
- “Pixabay.” n.d. Accessed February 5, 2021. <https://pixabay.com/>.
- “Qualitative Data Analysis Software | NVivo.” n.d. Accessed February 5, 2021. <https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home>.
- “Quantitative Social Science: An Introduction.” n.d. Quantitative Social Science. Accessed February 5, 2021. <http://qss.princeton.press/>.
- “R Programming A-Z: R For Data Science (Course & Exercises).” n.d. Udemy. Accessed February 5, 2021. <https://www.udemy.com/course/r-programming/>.

- “R: The R Project for Statistical Computing.” n.d. Accessed February 5, 2021. <https://www.r-project.org/>.
- “Senior Appointments and Corruption within Kabul City Police.” 2015. Integrity Watch Afghanistan. <https://iwaweb.org/blog/publication/justice-sector/senior-appointments-and-corruption-within-kabul-city-police/>.
- Sharma, Raghav, and Sayed Ikram Afzali. 2018. “Education Compromised? A Survey of Schools in 10 Provinces of Afghanistan.” Integrity Watch Afghanistan. <https://iwaweb.org/wp-content/uploads/2014/12/EDUCATION-COMPROMISED-A-SURVEY-OF-SCHOOLS-IN-10-PROVINCES-OF-AFGHANISTAN.pdf>.
- “Small SEO Tools.” n.d. Small SEO Tools. Accessed February 5, 2021. <https://smallseotools.com/>.
- “Swirl | Students.” n.d. Accessed February 5, 2021. <https://swirlstats.com/students.html>.
- “The Joint Program on Essential Services Package (Esp) for Women and Girls Subject to Violence in Vietnam: Baseline Assessment Report 2018.” 2019. UN Women. <https://asiapacific.unwomen.org/-/media/field%20office%20eseasia/docs/publications/2019/10/vn-baseline-assessment-report-esp-en.pdf?la=en&vs=5503>.
- “Write Your Best with Grammarly.” n.d. Accessed February 5, 2021. <https://www.grammarly.com/>.

## Appendix A: Checklist for creating a good TOC

1. Are all elements including resources, activities, and influential factors well defined?
2. Are there any gaps in the logical chain of events and are activities clearly connected to the specified program theory?
3. Are elements including major activities necessary and sufficient?
4. Are relationships plausible and consistent?
5. Is the theory plausible, feasible, and testable or are there other plausible pathways to achieving the program outcomes?
6. Is it realistic to assume that the program will result in the attainment of stated goals in a meaningful manner?

## Appendix B: Optimizing quasi-experimental approaches:

In cases where control groups are not chosen through randomized control trials, the following approaches can be useful for creating better quasi-control groups to compare the intervention group with. These approaches are categorized as upper- and lower-tier. Upper tier approaches include instrumental variables, regression discontinuity design, and difference-in-differences control for unobservable characteristics while lower-tier approaches including matching and non-equivalent comparisons do not.

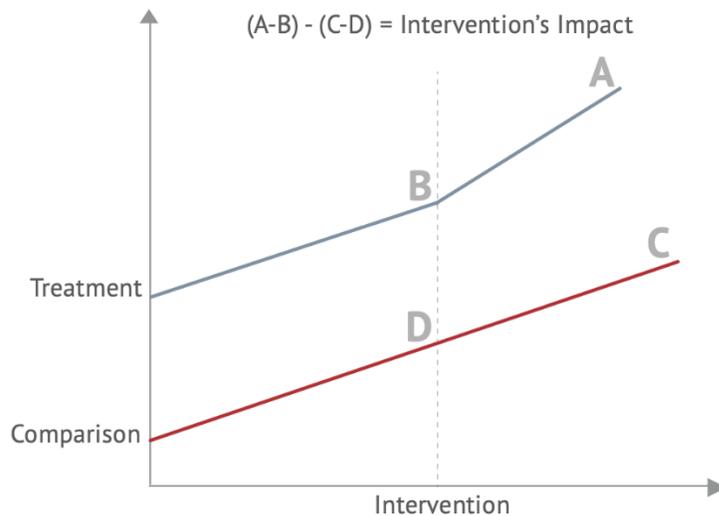
### **Quasi-experimental approaches: Upper-tier approaches that control for unobservable characteristics**

**Instrumental variables:** Researchers can measure instrumental variables that influence the independent variables instead. For example, since it is hard to measure the effects of income (independent) on happiness (dependent), researchers can use lottery wins as a more random instrumental variable to investigate the effect of income on happiness. It is easier to measure the instrumental variables that influence the independent variable because they are more free from biases such as selection, time effects and reverse causality. This usually happens when the Instrument Variable occurs randomly (like the lottery wins above) or is randomly assigned by the evaluator (by extra encouragement, for example). Instrumental variables need to affect individual's probability of participating in a program but cannot have a direct impact on participant's outcomes. Once instrumental variables are selected, create control and treatment groups that have different levels of this variable.

**Regression discontinuity design:** This design involves only measuring individuals who are just eligible (treatment) and just ineligible (control) for a program. Eligibility is determined using some quantifiable threshold criterion i.e. income level, age, grade point averages. Above and below the criterion are the just eligible and just ineligible groups that are effectively comparable because of the slight differences in their status due to random variation. For example, if a program admits individuals that score above 40%, the study would compare those who scored between 35-39% and those who scored between 40-44%. For this type of comparison, there needs to be enough cases clustered around criterion to enable an analysis. The criterion for eligibility also cannot be correlated with anything else that might explain outcomes. A disadvantages of this design is that it is more limited in external validity when it revolves around a small portion of cases around the cut-off mark.

**Difference in differences:** Use before and after data from a treatment and comparison group to compare changes in outcomes between them. The figure below indicates data points to measure in order to determine an intervention's impact. This design is useful when it's reasonable to assume that parallel trend assumptions hold.

Figure 18: Difference in Difference assessments



### Lower-tier quasi-experimental approaches that do not control for unobservable characteristics

**Matching:** A technique for matching each participant or unit of intervention (treatment) with a similar non-participant (control) based on observable traits. The match should have similar values for a set of confounding variables such as age, gender, income, education. It is harder to match if there are more variables that we attempt to match on and there might not always be an exact match for each participant.

**Propensity score matching** is a variation of matching estimates an individual's propensity to enroll in an intervention based on a set of observable characteristics that predict enrollment. Participants are matched with non-participants who have identical or near-identical scores.

Matching is useful when large data sets on participants and non-participants exist. It assumes that the treatment and control sub-samples are comparable because there are no unobservable characteristics predicting participation. For propensity score matching, there needs to be good

knowledge of factors that predict program participation. All variables that cause bias must be identified and included in matching algorithm.

Steps for matching:

1. Estimate model of program participation
2. Define region of common support and balancing tests
3. Match participants to nonparticipants
  - Nearest-neighbour matching: matched to unit with closest propensity score
  - Caliper or radius matching: threshold or tolerance on max score distance, allows matching with replacement
  - Stratification or interval matching
  - Kernel and local linear matching: partitions common support into different intervals: if common support criteria too narrow
4. Calculate Average treatment effect= mean difference in outcomes across these 2 groups
  - Advantages:
    1. Relies on degree to which observed charac drive program participation
    2. Does not require baseline data
    3. Good for demand-driven projects where cannot create control group because communities self-select into programs

Example of matching: The TRABAJAR II Project in Argentina.

This project was focused on providing employment at low wages in small social and economic infrastructure subprojects selected by community groups. The impact evaluation of the program was designed to assess whether the incomes of program participants were higher than they would have been had the program not been in place. The most commonly used methods to estimate household income without the intervention were not feasible in the case of the TRABAJAR program: no randomization had taken place to construct a control group to use in comparing the income of project beneficiaries; and no baseline survey was available, ruling out the possibility of conducting a before-and-after evaluation. The TRABAJAR evaluation instead used existing data to construct a comparison group by matching program participants to nonparticipants from the national population over a set of socioeconomic variables such as schooling, gender, housing, subjective perceptions of welfare, and membership in political parties and neighborhood associations by using a technique called propensity scoring. The study demonstrates resourceful use of existing national household survey data—the Encuesta de Desarrollo Social (EDS)—in generating the comparison group, combined with a smaller survey of TRABAJAR participants conducted specifically for the purposes of the evaluation. The

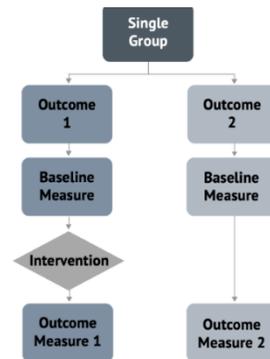
smaller survey was carefully designed so that it used the same questionnaire as the EDS and the same interview teams and was conducted at approximately the same time in order to successfully conduct the matching exercise. This technique was possible in the TRABAJAR case because a national household survey was being canvassed and the evaluators could take advantage of this survey to oversample TRABAJAR participants. The same interview teams were used for both the national and project surveys, resulting in efficiency gains in data collection

Baker, Judy L., and Bank World. *Evaluating the Impact of Development Projects on Poverty : A Handbook for Practitioners*, World Bank Publications, 2000. ProQuest Ebook Central, <http://ebookcentral.proquest.com/lib/oculcarleton-ebooks/detail.action?docID=714463>. Created from oculcarleton-ebooks on 2020-10-21 04:04:20.

**Non-Equivalent Comparison:**

This design compares change in outcome variable to changes in other similar variables between the treatment and control group. The variable used as a control is not affected by treatment but is measured on the same scale. The compared variable must be prone to all the same factors as the treatment variable. The two variations include pre and post comparison or Post only comparison. The post-only is a weaker design because differences between the intervention and comparison group may reflect differences in where they began rather than the effect of the intervention.

Figure 19: Non-Equivalent Comparison



For more information on how to conduct quasi-experimental research designs, refer to [Measuring impact by design- Impact Canada](#)

## Appendix C: Alternatives to establishing causation

Establishing counterfactuals usually works when causal factors are independent. Here are three other approaches to draw causation based on multiple causes and outcomes, each with its own requirements, strengths, and weaknesses.

- **Regularity frameworks** infers causation based on the high frequency of association between cause and effect from a high numbers of diverse cases. It works best when causal factors are independent and is weaker at dealing with contextualising causation.
- **Multiple causation** depends on combination of causes that lead to an effect. Drawing from a sufficient number of cases with comparable characteristics, this approach is good at dealing with limited complexity and interdependence.
- **Generative causation** depends on identifying the mechanisms that explain effects. This process-based approach requires one case with good access to data sources as well as supporting theories. It is weaker at external validity.

*Table 12: Strength of Causal Inference*

	Narrow scoped- strong	Broad scoped- weaker
Advantages and disadvantages	clinch the conclusion but are narrow in their range of application	vouch for the conclusion but are broad in their range of application
methods	Randomized Controlled Trials	QCA: qual comparative analysis

## Appendix D: Cost-effectiveness analysis

**Cost-effectiveness analysis** is a technique that relates total costs to some measure or measures of program effectiveness. **Cost-benefit analysis** takes that process one step further and seeks to place dollar values on all (or most) costs and benefits to calculate net benefits

### Framework for analysis

1. Identify all known benefits and costs. Be careful not to confuse benefits and costs with transfers which alter the distribution of resources within society. Distinguish:
  - **direct vs indirect** benefits and costs. Indirect are byproducts, multipliers, spillovers, investment effects
  - **Marginal benefits** and costs: marginal = incremental generated by one more unit of output
  - **Fixed vs variable** costs: variable depend on size of programs
  - **Tangible vs intangible**: tangible can readily be identified in unit terms
  - **geographical scope of analysis**- spillovers into neighbouring areas that have political consequences
2. Add dollar value to costs which could include cost of capital and resources as well as indirect costs such as
  - Overhead
  - Costs to private sector
  - Cost to participant and volunteers
  - Sunk costs: investments prior to the program or project
3. Add dollar value to benefits while taking into account the following indirect benefits:
  - Cost avoidance
  - Time saved
  - Valuing lives
  - Increased productivity
  - Land values increases
  - Taxes
4. Compare the benefits and the cost while taking into account these common issues and problems with cost-benefit analysis
  - Equity concerns: weight benefits and costs accordingly
  - Sensitivity of the analysis to particular assumptions; that is, what is the probability that those particular assumptions will occur?
  - Spillovers and unintended effects of government actions

# Appendix E: Focus Group Consent Form Example



THE UNIVERSITY of EDINBURGH  
School of Geosciences

Here is an outline of the information you should consider including on a 'Focus group Consent' form. This template is not prescriptive and it is provided to act only as a guide for your research project requirements.

## Focus Group Consent Form

Research project title:

Research investigator:

I agree to participate in the (name of focus group) carried out by (name of researcher) of the University of Edinburgh, to aid with the research of (name research project).

- I have read the information sheet related to the (name the research project) and understand the aims of the project.
- I am aware of the topics to be discussed in the focus group.
- I am fully aware that I will remain anonymous throughout data reported and that I have the right to leave the focus group at any point.
- I am fully aware that data collected will be stored securely, safely and in accordance with Data Collection Act (1998).
- I am fully aware that I am not obliged to answer any question, but that I do so at my own free will.
- I agree to have the focus group recorded (video or dictaphone), so it can be transcribed after the focus group is held. I am aware that I have the right to edit the transcript of the Focus Group once it has been completed.
- I am aware that I can make any reasonable changes to this consent form.

---

**Printed Name**

---

**Participants Signature**

---

**Date**

---

**Researchers Signature**

---

**Date**

105

Contact Information

This research has been reviewed and approved by the Edinburgh University Research Ethics Board. If you have any further questions or concerns about this study, please contact:

Name of researcher

Full address

Tel:

E-mail:

You can also contact (Researcher's name) supervisor:

Name of researcher

Full address

Tel:

E-mail:

**What if I have concerns about this research?**

If you are worried about this research, or if you are concerned about how it is being conducted, you can contact the Chair of the GeoScience Ethics Committee, University of Edinburgh, Drummond St, Edinburgh, EH8 9XP (or email at [ethics@geos.ed.ac.uk](mailto:ethics@geos.ed.ac.uk)).

Add names of any associated funding bodies and their logos

# Appendix F: Template for an Evaluation Term of Reference

## Template:

- Background and context
- Evaluation purpose
- Evaluation objectives
- Criteria and key questions
- Methodology
- Evaluation products expected
- Ethics
- Team composition and required competencies
- Management and implementation arrangements and challenges
- Time frame
- Budget that takes into account the scope, complexity and time commitments of evaluation.
- TOR annexes
  - Intervention results framework and Theory of Change
  - Key stakeholders and partners
  - Docs to be reviewed and consulted
  - Code of conduct forms