
Empirical Research and the Development and Assessment of Transitional Justice Mechanisms

Phuong Pham* and Patrick Vinck†

Abstract¹

In this paper, we demonstrate different methods of empirical research available to transitional justice practitioners and scholars. Guidelines on how to conduct research in the field of transitional justice are outlined on the basis of the principles of monitoring and evaluation for decision making, program development and policy development. We argue that such methods offer policy makers a systematic way to consult a population and construct a comprehensive view of what this population has experienced (e.g., violations), what they know (e.g., knowledge of various transitional justice mechanisms), what they believe in (e.g., definition of justice) or what they need (e.g., accountability). Such knowledge is essential for developing effective evidence-based transitional justice programs.

Introduction

Over the last two decades, societies have increasingly embraced the enforcement of international human rights norms and demanded greater accountability when confronted with intrastate mass violence, war or transition from repressive rule.² Various transitional justice mechanisms³ have been implemented in such countries as South Africa, the states of the former Yugoslavia, East Timor, Iraq and

* Phuong Pham, PhD is Adjunct Assistant Professor at the Payson Center for International Development and Technology Transfer, Tulane University and Senior Research Fellow at the Human Rights Center, University of California, Berkeley, USA. Email: ppham1@tulane.edu

† Patrick Vinck, PhD is Director of the Berkeley-Tulane Initiative on Vulnerable Populations, Human Rights Center, University of California, Berkeley, USA. Email: pvinck@berkeley.edu

¹ The authors are indebted to Eric Stover and Harvey Weinstein of the Human Rights Center, University of California, Berkeley for their vision of integrating social science research methods in the field of human rights and transitional justice. We are also grateful to the Human Rights Center and International Center of Transitional Justice for involving us in their programs and allowing us the freedom to develop our research with them. We also send our gratitude, respect and admiration to Drs William Bertrand and Eamon Kelly of the Payson Center for International Development and Technology Transfer and Drs Roger Sherwin and Janet Rice of the School of Public Health and Tropical Medicine, Tulane University for helping us strengthen the applications of empirical research in complex settings. Finally, we thank Eric Stover, Eamon Kelly and Joshua Rodd for their input on the draft manuscript.

² Roy Licklider, 'Obstacles to Peace Settlement,' in *Turbulent Peace: The Challenges of Managing International Conflict*, ed. Chester A. Crocker, Fen O. Hampson and Pamela Aall (Washington, D.C.: United States Institute of Peace Press, 2001).

³ The International Center for Transitional Justice (ICTJ) defines transitional justice as 'a range of approaches that societies undertake to reckon with legacies of widespread or systematic human

Rwanda among others. Despite an emerging consensus that transitional justice policies must take into account the local context, including culture, economic factors and political influences,⁴ local communities are rarely consulted to inform the design, implementation and evaluation of such policies. Local population participation in the decision-making process, planning and implementation has been identified as the most central feature of successful development programs.⁵ At a basic level, participation can be achieved through consultation. Empirical research⁶ can provide precise assessments of communities' needs and perceptions of and attitudes towards peace and justice, as well as systematic and rigorous measurement of the potential and actual impacts of transitional justice mechanisms. Furthermore, the analysis and comparison of empirical data allows researchers to test hypotheses and develop a better understanding of the factors that affect how individuals and communities confront past atrocities. Such knowledge can contribute to theory building and is essential for transitional justice advocates and practitioners to effectively anticipate and address the need for and the impact of policies.⁷ Several collective efforts, including three key conferences,⁸ have been held to discuss how empirical research can contribute to the field of human rights and transitional justice.

Stover and Weinstein note that:⁹

Unfortunately, as we found in Rwanda and the former Yugoslavia, most international aid agencies and their national counterparts forgo collecting and analyzing population

rights abuse as they move from a period of violent conflict or oppression towards peace, democracy, the rule of law, and respect for individual and collective rights.' ICTJ, 'What is Transitional Justice?' <http://www.ictj.org>. Transitional justice mechanisms include criminal trials (national, international or hybrid), truth commissions, reparations measures for victims and communities, policies to remove past offenders from office (e.g., de-Nazification programs in post-war Germany, lustration programs in Eastern Europe and de-Ba'athification in Iraq), and reform of institutions to conform to international standards and to address root causes of inequalities.

- ⁴ United Nations Secretary General, *The Rule of Law and Transitional Justice in Conflict and Post-conflict Societies, Report of the Secretary General, S/2004/616* (United Nations Security Council, 2004).
- ⁵ John M. Cohen and Norman Uphoff, 'Participation's Place in Rural Development: Seeking Clarity through Specificity,' *World Development* 8(3) (1980): 213–35.
- ⁶ We define empirical research as the systematic data collection of information or knowledge using scientific approaches. In such a positivist approach, rigorous attention is paid to probability sampling, measurement and control of bias. Statistical analysis is used to demonstrate relationships between variables.
- ⁷ In their arguments on the use of social science theories to explain ethno-political conflicts, Gurr and Harff note that knowledge is essential for advocates, program developers, scholars and policy makers to understand, anticipate and respond to ethnic conflict. Ted Robert Gurr and Barbara Harff, *Ethnic Conflict in World Politics* (Boulder: Westview Press, 1994), 98–103. We extend this proposition to encourage the use of monitoring and evaluation research methods to inform the selection of transitional justice mechanisms and the monitoring of their impact in emerging societies.
- ⁸ The three conferences are: (i) 'Empirical Research Methodologies in Transitional Justice,' in Stellenbosch, South Africa in November 2002; (ii) 'Measuring Progress, Assessing Impact,' at the Carr Center for Human Rights Policy in May 2005; and (iii) 'Beyond the Toolkit: Rethinking the Paradigm of Transitional Justice,' held in Bellagio, Italy in October 2006.
- ⁹ Eric Stover and Harvey M. Weinstein, 'Conclusion: A Common Objective, a Universe of Alternatives,' in *My Neighbor, My Enemy: Justice and Community in the Aftermath of Mass Atrocity*, ed. Eric Stover and Harvey M. Weinstein (Cambridge, UK: Cambridge University Press, 2005), 326.

data prior to launching social reconstruction projects because they lack the appropriate expertise, or consider data collection too time-consuming, or fear it will cause friction among former belligerents. As a result, millions of dollars were wasted in these countries on ill-conceived projects that failed to meet the needs of those affected by the violence.

Building on this premise, this paper presents a framework for evidence-based transitional justice based on the principle and theory of monitoring and evaluation for decision making, program development and policy development. The paper targets transitional justice decision makers, program development staff and researchers interested in applying quantitative research to the field of transitional justice. We use the term 'evidence-based transitional justice' to describe the process of using empirically based data to inform the development, implementation and evaluation of transitional justice mechanisms. The evidence-based approach refers to the systematic and scientific monitoring and evaluation of the context and impact of interventions. It is a process of gathering, critically examining and applying information to develop, implement and evaluate programs and policies in order to improve performance, outcome and the use of resources to achieve that outcome.¹⁰ Evidence-based transitional justice consists of the initiation of policies and programs based on evidence derived from the best available data.

In this paper, we offer a conceptual framework to guide the development of such research. We define the tools and terminologies used in monitoring and evaluation research, the types of evaluation and research questions and the life-cycle of empirical research. Our aim is not to advocate for one specific methodological approach over another, but rather to discuss how different research methods are useful depending on the context and type of research questions. Where possible, we illustrate the discussion using our own experience with population-based empirical research to study attitudes toward justice and peace in war-torn and post-conflict countries such as Iraq,¹¹ Rwanda¹² and Uganda.¹³ While much of the paper is targeted at a general audience, it is informed by the authors' experiences of doing such research as foreign researchers in these countries. Its recommendations and lessons learned are therefore written from that specific perspective.

¹⁰ Obioha C. Ukoumunne et al., 'Evaluation of Health Interventions at Area and Organization Level,' *British Medical Journal* 319 (1999): 376–379.

¹¹ Human Rights Center and International Center for Transitional Justice, *Iraqi Voices, a Population based Survey of Attitudes toward Peace and Justice* (Human Rights Center, University of California Berkeley and International Center for Transitional Justice, 2004).

¹² Phuong Pham, Harvey M. Weinstein and Timothy Longman, 'Trauma and PTSD: Their Implication for Attitudes towards Justice and Reconciliation,' *Journal of the American Medical Association* 292(5) (2004): 602–612.

¹³ Phuong Pham et al., *Forgotten Voices, Attitudes toward Peace, Justice and Reconciliation in Northern Uganda* (Human Rights Center, University of California Berkeley and International Center for Transitional Justice, 2005).

Approach to Qualitative and Quantitative Data Collection

Human rights and transitional justice researchers often debate the value of either qualitative or quantitative methods, and sometimes of both of them. A common mistake is to assert that either method is intrinsically superior to the other. The two methods serve different purposes and should be seen as complementary. Research is not a matter of collecting qualitative or quantitative data but, rather, involves the strategic collection of data that will best suit the objectives of the research and assist in evidence-based decision making for program and policy development.¹⁴

Qualitative research methods involve collecting textual data through interviews and/or observation of individuals or groups of people. Some common methods are key informant interviews, focus group discussions, participant observation and direct observation. Qualitative research offers a depth and richness of response that illuminates the dynamics of the process under study. It can also identify more complex themes such as social identity and the power of religious beliefs, which is something a structured questionnaire survey cannot do. When qualitative research is used in conjunction with quantitative research, it can help in: developing quantitative research questions, measures and instruments; generating hypotheses; and explaining quantitative findings. However, because qualitative research typically does not involve random sampling of respondents and usually focuses on a small number of individuals or a community, it is not ideal for answering questions about the magnitude of phenomena or for generalizing findings to the population as a whole. Non-random samples, with the possibility of significant bias, may lead the researcher to reach conclusions that do not reflect a more widely-held attitude, belief or behavior.

In contrast, quantitative research methods entail the collection and/or analysis of data that can be measured numerically. Generally this is done through the use of structured interview instruments with numeric coding of responses or the analysis of secondary data, or data that has been collected for other purposes (for example, analyzing the World Bank or United Nations Development Program development indicators). The advantages of using quantitative data are that the results can theoretically be replicated through repetition of the same study protocol; measurements of precision and error can be estimated through statistical techniques; systematic errors (i.e., bias) can be detected; a large number of individuals can be interviewed (larger sample size); analyses usually take less time; and complex concepts and associations can be tested statistically and hypotheses examined. In the example of transitional justice, quantitative data can be used to measure the frequency of support for various mechanisms and even to establish association of these attitudes with predictive factors such as exposure to trauma,

¹⁴ Donald T. Campbell, 'Degree of Freedom and Case Study,' *Comparative Political Studies* 8 (1975): 178–193.

whereas qualitative data are best placed to describe what people understand by keywords such as 'justice' or 'reconciliation.' Quantitative data are limited by the kinds of questions asked and by the definitions assigned to the concepts involved. Questions are usually structured and hence offer little freedom to adjust to participants' responses. For this reason it is critical to know the cultural and social context in order to ask the right questions in the right way, use the correct language and avoid the pitfall of making assumptions about what certain behaviors or concepts mean in a specific culture or subculture.

In Iraq, the authors conducted a qualitative assessment of attitudes towards peace and justice in collaboration with colleagues from the Human Rights Center at the University of California, Berkeley and the International Center for Transitional Justice.¹⁵ A total of 38 key informant interviews and 49 focus groups with Iraqis from a variety of ethnic, religious and political backgrounds were conducted in July 2003, which was four months after the fall of Baghdad and shortly before the insurgency strengthened. The qualitative approach was chosen because little information was available to inform the design of a quantitative survey and there were obvious security concerns involved in randomly selecting respondents.

In Uganda, however, a quantitative approach was chosen to examine attitudes towards peace, justice and reconciliation in northern Uganda.¹⁶ Population-based data were collected from 2,550 individuals in the region. The quantitative approach was chosen because it offered a systematic, unbiased way to assess the affected population's expectations for both peace and justice and provide basic statistics to inform the debate taking place between proponents of the International Criminal Court (ICC) and proponents of amnesty and/or local justice initiatives.

Type of Research: The Monitoring and Evaluation Framework

From the moment it is conceptualized to the moment of its completion, a transitional justice mechanism is expected to bring about changes, such as establishing the truth, achieving retributive, reparative or restorative justice and contributing to peace and reconciliation.¹⁷ This raises the question of the timing of research: at what stage in the development and implementation of transitional justice mechanisms should empirical research be carried out? To answer this question we present a specialized field of research referred to as program monitoring and evaluation (M & E). M & E is not limited to assessment of policies and programs already in place, but also includes the collection of empirical evidence to inform decision makers at the planning stage through, for example, a thorough needs assessment.

¹⁵ Human Rights Center and International Center for Transitional Justice, *supra* n 11.

¹⁶ Phuong Pham et al., *supra* n 13.

¹⁷ Whether or not transitional justice can achieve such goals is debated; see, for example, Laurel E. Fletcher and Harvey M. Weinstein, 'Violence and Social Repair: Rethinking the Contribution of Justice to Reconciliation,' *Human Rights Quarterly* 24 (2002): 573–639.

Three stages of program and policy development can be broadly identified: (i) conceptualization and planning; (ii) implementation; and (iii) completion.¹⁸ Each stage poses different questions and challenges for decision makers. Accordingly, empirical research can be timed to answer those questions and challenges. At the conceptualization and planning stage, the central question for decision makers is: ‘What justice for whom and when?’ During implementation, the main question is: ‘Is the project working?’ Finally, upon completion, the question is: ‘What is the impact?’ Central to the M & E framework is the idea that empirical research is more than a snapshot in time and, instead, that it feeds into a dynamic decision-making, program-development and implementation process. For example, new constraints and needs may appear at the implementation stage and require additional evidence-based planning and policy or program changes. Accordingly, the M & E framework of empirical research distinguishes between three types of research, each with different timing and objectives: (i) formative evaluation; (ii) process monitoring and evaluation; and (iii) outcome monitoring and evaluation.

Formative Evaluation

Formative evaluation aims to answer the first question we raised: ‘What justice for whom and when?’ It is a type of needs assessment whereby data are collected at the population level or from key stakeholders. These data can be used for purely descriptive purposes or for examining the factors associated with the needs and opinions expressed by the respondents. Formative evaluation aims to answer questions about what happened in the past with regards to human rights violations (e.g., patterns of abuses); the attitudes towards justice, peace and reconciliation; who should be held accountable and through what means; and the nature of the transitional justice mechanisms that are currently in place. Further questions should investigate the population’s knowledge and perceptions of and attitudes towards the transitional justice mechanisms and the time that transitional justice policies should be implemented. This type of evaluation should take place during the policy conceptualization stage to inform decisions concerning the transitional justice mechanisms that should be pursued and when to pursue them. Formative evaluation provides a means to gauge what the population believes to be the priorities for transitional justice, as well as a platform for respondents and affected communities to contribute to the transitional justice and rebuilding processes.

The research we conducted in Uganda, for example, found that peace and a return to ‘normal life’ was a clear priority among the respondents, but this did not necessarily include forgiving and forgetting what had happened. About three-quarters (76%) of the respondents said that those responsible for abuses should be held accountable for their actions. However, 71 percent stated that they would

¹⁸ See, Peter H. Rossi, Mark W. Lipsey and Howard E. Freeman, *Evaluation: A Systematic Approach* (Newbury Park, CA: Sage Publications, 1999).

accept amnesty if it were the only road to peace. Given the choice, 54 percent would have preferred peace with trials over peace with amnesty (46%).

Process Monitoring and Evaluation

While formative evaluation informs the planning stage, process monitoring and evaluation is designed to answer the question: 'Is the transitional justice mechanism being effectively implemented?' Process monitoring refers to the descriptive process of collecting data about the characteristics of the population being served, the services provided and the resources used to deliver those services. Process evaluation is a normative process in which the responses to these questions are systematically compared to goals, objectives and standard procedures: 'Is the intervention being delivered to the intended beneficiaries and, if not, why not?' Specifically, process evaluation provides answers to questions as to whether the intervention was implemented as intended, whether it reached the intended audience and what barriers constituents experienced in accessing the intervention. The information collected assists program managers and program staff in ensuring that the appropriate activities are being implemented effectively and determining whether improvements or changes need to be made. Process monitoring and evaluation of a reparations program such as the International Criminal Court's Victim Trust Fund would answer questions about who receives compensation and the kind of compensation provided, whether eligible victims are receiving their compensation and, should this not be the case, the reasons for this. In the case of the International Criminal Tribunal for Rwanda, process monitoring and evaluation could answer questions related to court proceedings, such as: the court's success in identifying all the perpetrators that need to be prosecuted; whether trials take place on time and within budget; whether witnesses are provided with adequate protection; and whether the transcriptions and translations of the court proceedings are done according to regulations.

Outcome Monitoring and Evaluation

The third general question, corresponding to the completion of any transitional justice mechanism, is: 'What is the impact?' Transitional justice interventions are utilized to achieve different goals in different contexts. Amongst their primary goals, however, are deterring and preventing future conflicts, the discovery and publicizing of the truth about atrocities, the punishment of perpetrators, responding to the needs of victims, promoting the rule of law in emerging democracies and promoting reconciliation. Whether or not a specific policy has achieved any or all of these goals can be assessed through outcome monitoring and evaluation. Outcome monitoring first determines whether or not the expected outcomes were realized (e.g., peace was achieved), regardless of the role of the program or policy in achieving those outcomes. Outcome evaluation then assesses whether or not a causal link exists – such as, for example, whether a truth commission contributed to a sustainable peace.

Research Lifecycle

Once the type of evaluation and the general research questions have been identified, the following five steps can be used to guide the research: (i) conceptualizing the study; (ii) designing the research protocol; (iii) implementing the study; (iv) data analysis and reporting; and (v) dissemination and follow-up. The framework that we describe relies on methods of social science research and statistics and the authors' experiences in studying transitional justice in challenging environments such as (post-) conflict countries. We offer this guide to assist field-based organizations in securing data that will provide an accurate assessment of the questions relevant to both policy and intervention.

Step 1: Conceptualizing the Study

Typically, researchers have preliminary ideas on why and how they want to conduct a study. Formal study conceptualization is often overlooked because it is a time-consuming and challenging process. Too often, good research ideas have failed to develop into good research projects because insufficient attention was paid to the conceptualization phase. The result is, in general, a lack of depth in the data and possibly a lack of accuracy.¹⁹ A thorough conceptualization of the study is needed to collect accurate data, especially in an environment that is rich and complex in culture and tradition. This can be achieved through gathering as much information as possible through a literature review, consultations with key stakeholders and field feasibility assessments. One effective tool is appropriately called 'one thousand questions.'

According to this tool, when conceptualizing a study, researchers should compose and ask themselves and others at least one thousand questions. These 'one thousand questions' are best addressed through a multidisciplinary approach with the aim of providing cross-discipline insights and a variety of points of view for examining the problem at hand. By the end of this process, all there is to know about the context, content and issues at hand should be known and formally documented. Complex concepts such as peace and justice should be broken into simpler, measurable concepts.

Our experience in northern Uganda provides a concrete example of the use of the 'one thousand questions' tool. The impetus for our research on attitudes toward peace, justice and reconciliation in northern Uganda was based on partial information from reports about a debate regarding the need for accountability mechanisms in the region. The first step in the research was to spell out possible

¹⁹ For example, in northern Uganda 40 percent of the respondents reported being abducted at some point during the conflict (Phuong Pham et al., *supra* n 13). This was more than had been anticipated on the basis of information available from previous studies. Possibly, many people considered short encounters with the enemy as abduction. The questions about people's exposure to traumatic events did not distinguish between types of abduction because we failed to ask ourselves 'What does it mean to be abducted?' We lost some depth by failing to distinguish between short- and long-term abduction. Such difficulties cannot be avoided; there are always some unasked questions that become obvious only during the data-analysis phase.

questions to anchor the study, whether related to theory or to the context of northern Uganda. There were four initial underlying questions: (i) Who are the victims? (ii) Which transitional justice mechanisms are being considered? (iii) What are the options for peace? and (iv) What is the context in which the various options are being considered? The basic principle of the ‘one thousand questions’ approach is that each question should raise another set of questions. For example, ‘Who are the victims?’ raises a series of new questions, such as questions on: what violations victims have experienced; where they are (e.g., in displaced camps); how others have measured the concept of victimhood (e.g., psychological trauma exposure, self-perception); socio-economic profiles; whether some groups have been targeted for abuses more than others; and the relationship of the victims to the offenders (e.g., is this an ethnic or a religious conflict). Again, each of these questions raises yet another set of questions. When asking what victims have experienced, one must also ask oneself (or the literature, or local experts) about the types of violence committed in northern Uganda. If the answer is abduction then the next question has to discover what abduction means in the northern Ugandan context: the period of time people are abducted for; whether abduction entails being forced to commit crimes; and if evidence exists that abduction affects behaviors and/or attitudes.

Answers to some questions will inform the content of the research, while others will inform the feasibility of the study and the methodology to be used. One of the questions mentioned above asked where victims are; building on the response to this question, we can ask where and from whom the researchers should collect data. Other questions that naturally follow are: ‘Are there security concerns?’; ‘Is the area accessible?’; ‘What languages do people speak?’; ‘What is the general level of education (as literacy can be a factor in deciding on specific approaches)?’; ‘How much would the proposed mechanism cost?’; ‘What resources are available?’; and ‘What objectives can be achieved with those resources?’

It is likely that not all of the questions posed will have an answer; this is the point of conducting the research and gathering new information that is not yet available. The ‘one thousand questions’ approach enables the researcher to gain a deeper understanding of the situation and be better equipped to conduct informed research that pays close attention to cultural and historical context. The researchers will also be able to make informed decisions about the costs and benefits of various research options (e.g., deciding whether it is advisable to go into isolated locations along the border of Sudan, knowing that the security situation there is precarious). Research studies always entail a compromise between the scope of the research, the resources available and feasibility. They do not however always need to be expensive endeavors, and can be conducted with limited resources where necessary.

Typically, answering the one thousand questions requires an in-depth literature review and the involvement of experts and people knowledgeable of the local cultures and context. At the very least, one feasibility mission should be conducted before the research methods and data-collection protocol are developed. This is

crucial for working during or in the immediate aftermath of a disaster or outbreak of violence. Information gathered during the feasibility mission not only helps to answer all of the questions set out above, it also aids in determining whether or not: security conditions will permit dispatching a team on the ground; local partners are available; the population is willing to talk openly about their experiences; and the community and local authorities accept and permit the research to be conducted. Often, complex emergencies or post-conflict settings limit the time available in the field and the use of prolonged data-collection methods such as ethnographic methods.

The conceptualization phase forms the backbone of the study. By the end of the conceptualization phase, the researcher is expected to: (i) define a set of clear research questions and objectives or hypotheses; (ii) develop a conceptual framework of the research components (See Figure 1);²⁰ (iii) list key concepts, variables and indicators; and (iv) estimate the necessary financial and human resources needed for the research. In addition, researchers need to be cognizant and assess the ethical issues that are integral to research involving human subjects. Where possible, all studies should be reviewed by Human Subjects Committees in the countries involved.²¹

Step 2: Research Planning

Once the study is conceptualized, practical planning of the research begins. Planning includes three components: (i) the structural design (research-protocol development, questionnaire design, back-translation, assessment of potential bias); (ii) the implementation plan (logistics, timeline, budget, feasibility re-assessment); and (iii) the piloting and training (small-scale implementation, quality control and training). During the physical design, the researcher must determine the research protocol that will be used by defining what types of data will be collected, what type of study design will be used and, if appropriate, how many people should be interviewed. The number of participants in the study may be a critical factor deciding on the use statistical methods to assess the significance of findings and/or the associations of factors.

Study Design

In general, research designs (i.e., data collection ‘templates’) can be classified into two major categories – observational and experimental. Each of these study

²⁰ The conceptual framework is a schematic representation of the relation between the different components measured, which pictorially represents the research question and hypothesis.

²¹ Such a committee may not exist or be functional in (post-) conflict countries. In such cases, researchers should follow regulations and ethical guidelines for the conduct of research involving human subjects as outlined in: (i) The Nuremberg code (*Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law No. 10, Vol. 2, 181–182*. Washington, D.C.: U.S. Government Printing Office, 1949); (ii) The Declaration of Helsinki (Adopted by the 18th World Medical Association General Assembly, June 1964 and subsequent amendment); and (iii) The Belmont Report (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, April 18, 1979).

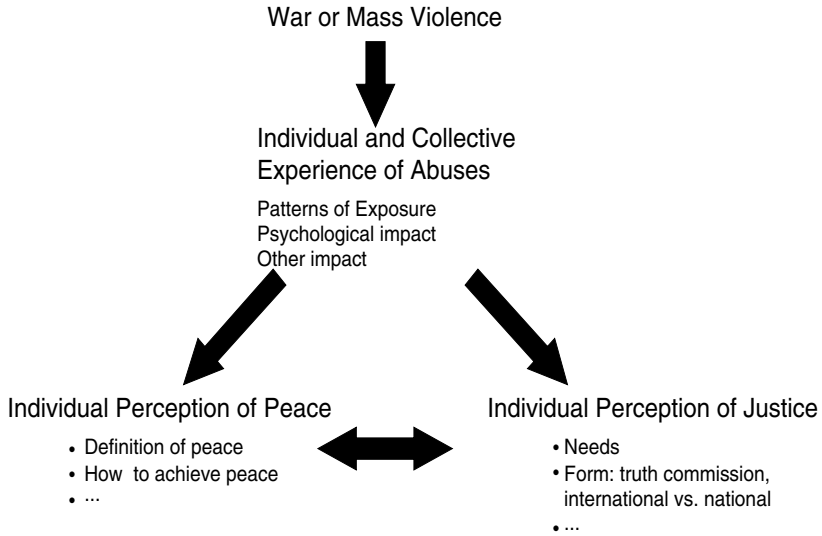


Fig. 1 Example of conceptual framework: victim-centered model of peace and justice²²

designs has its strengths and weaknesses, and the type of study design appropriate for research depends on the type of research questions being asked, the resources available and the context of the research. Observational study designs are usually used for descriptive, normative and correlative research questions. Observational studies cannot answer questions on causation, but they can help establish propositions and generate hypotheses. Examples of two types of observational designs that may be used in the field are case studies and cross-sectional surveys. Case studies build on various sources of evidence to examine a single phenomenon, event or policy within its context.²³ David Cohen, for example, conducted a case study of the UN-sponsored serious crimes process in East Timor.²⁴ Cross-sectional surveys use quantitative data collection to assess a snapshot of the population or a subset of the population predictor (or intervention) and outcome measures at a given time. The Uganda study introduced above²⁵ is an example of a cross-sectional survey designed to answer descriptive and correlative questions (e.g., what

²² This framework, developed by Patrick Vinck, proposes that exposure to war or mass violence (victimization) influences how individuals perceive peace and how individuals perceive justice. The framework also proposes that how individuals perceive peace is associated with how they perceive justice.

²³ Robert K. Yin, 'Case Study Research. Design and Methods Third Edition,' *Applied Social Research Method Series Vol. 5* (Newbury Park, CA: Sage Publications, 2002).

²⁴ David Cohen, 'Indifference and Accountability: The United Nations and the Politics of International Justice in East Timor,' *East-West Center Special Reports No. 9* (Honolulu: East-West Center, June 2006).

²⁵ Phuong Pham et al., *supra* n 13.

is the level of support for trials or for amnesty? Is there a relationship between exposure to violence (predictor) and attitude towards trials (outcome)?

Experimental study designs are often used for impact evaluation and causal study. There are three general types of experimental studies – quasi-experimental, randomized controlled and batch-randomized trial. The key concept of experimental studies is that they compare variables measured within a group exposed to a specific intervention to the same variables measured within a group not exposed to that intervention. This method allows the researcher to isolate the impact of the intervention from other factors.

Sampling Strategy

The final step in the research protocol is establishing the sampling strategy. Data are rarely collected from an entire population other than for the purposes of a census. Instead, most research relies on sampling – the process of selecting a number of subjects from all the subjects in a particular group or ‘universe.’²⁶ Deciding who to interview or where to collect data from can be a challenging task. The two general types of sampling are probability and non-probability sampling. In most cases, quantitative research relies on probability sampling and qualitative research on non-probability sampling.

Probability sampling ensures that the selected respondents are representative of the population under study; this is the most challenging type of sampling because it requires that every individual entity within the targeted population has a certain probability of being selected that is known to the researcher.²⁷ In northern Uganda, several different options for sampling were available to us. In an ideal case, a list of all the individual entities that can be sampled (e.g., list of all households, adults, etc.) would be available. With such a list, known as a ‘sampling frame,’ individuals can be selected randomly (simple random sample) or systematically (e.g., one out of every ten units is selected) from it. However, in Uganda, like in many developing countries, post-conflict or disaster settings, comprehensive sampling frames were not readily available. An alternative, therefore, is to use cluster sampling in which larger units are first sampled, then all or a sample of individual entities within the cluster are selected; this is called ‘multistage cluster random sampling.’ In Uganda, camps or villages were first randomly selected (first stage) and then within those clusters individuals were randomly selected (second stage). In this way the researcher does not need to have a list of all individuals in every village or camp, but only a list of the individuals in the selected sites. There is an increase in sampling error²⁸ with this method because of the homogeneity of

²⁶ John Last, *A Dictionary of Epidemiology* (New York: Oxford University Press, 2000), 196.

²⁷ For more information, see, Paul S. Levy and Stanley Lemeshow, *Sampling of Populations, Textbook and Solutions Manual: Methods and Application* (New York: Wiley Interscience, 2003).

²⁸ Measures on a subset of the population (sample) always differ from the true value for the entire population. Sampling error, also known as the ‘standard error,’ occurs because not every individual in the population is measured. The importance of the difference between the sample and the true

the individual entities within the same cluster, and this needs to be adjusted for through the use of a larger sample size.

Another type of probability sampling is stratified random sampling. In stratified sampling, sampling units are grouped (stratified) on the basis of certain criteria such as gender and then sampling occurs within each group. For example, to ensure that gender is equitable in a sample, a researcher might sample 100 people from the list of women and 100 people from the list of men.

Unlike probability sampling, non-probability sampling has few formal restrictions and as a result is logistically easier to implement than probability sampling. To be truly representative of the population of northern Uganda, a random sample of individuals regardless of any of their characteristics (e.g., location, ethnicity, religion) is needed, as described above. However, we could decide to facilitate the process by going to a victim-counseling center and conducting interviews with all the people we found there. This type of sampling is called 'convenience sampling.' It is considered a type of non-probability sampling because people who do not go to the center have no chance of being interviewed. Inevitably, using non-probability sampling results in bias. In our example of the victim-counseling center, we are aware that only a small proportion of victims, and an even smaller proportion of the overall population, goes to counseling centers. People who have no transportation or those with little or no exposure are unlikely to be found at the counseling center. In other words, the people interviewed at a counseling center do not represent the overall population, but only a subset with specific characteristics. This in turn may affect the results collected (e.g., those at a counseling center may be more likely to have experienced psychological trauma or be more likely to seek revenge). Therefore, in most cases, a random sampling technique is needed to assure that generalizations can be made and that bias is kept to a minimum for a quantitative study.

Data-collection Instrument

Considerable time should be devoted during the planning phase to the development of the data-collection instrument to ensure that the right questions are asked in the right way. The questionnaire must be consistent with the research objectives or hypotheses, the conceptual framework and the list of key concepts, variables and indicators established in the conceptualization phase. Prior ethnographic research and the 'one thousand questions' are critical in informing the type and content of the questionnaire. Experts in the local context or in the subject at hand must review the questionnaire to ensure that concepts are accurately measured and the instrument is sufficiently culturally sensitive for collecting quality data.

population depends, among other things, on the sample size and the selection design. Cluster sample results in 'pockets' of observation as opposed to simple random sample. Because of that clustering effect, the sampling error increases.

When respondents do not speak the language in which the research is developed, translation is necessary. Typically, the original version is translated into the local language, preferably by a group of local experts with both content and local language knowledge. The local translation should then be independently translated back into the original version (back-translation) so that any discrepancies between the original version and the re-translated version can be addressed. For example, translation of a questionnaire in Uganda showed that amnesty, forgiveness and reconciliation are similar concepts in the Acholi, Langi and Ateso languages.²⁹ The Acholi word 'Timo-kica' can mean either 'amnesty,' 'forgiveness' or 'reconciliation.' Extensive consultation was necessary to carefully select qualifier terms such as 'legal pardoning' and 'living together in harmony and unity' that would allow us to distinguish between the three concepts.

Once finalized, the questionnaire must be piloted. This stage is often overlooked, but the small-scale implementation of the entire study provides an opportunity to identify problems and to revise the protocol and questionnaire accordingly. Questions may be rephrased if respondents have problems answering. When response options are provided, the pilot phase is a good opportunity to ensure that those options exhaustively represent actual responses. The pilot phase is also a vital part of interviewer training because it allows for the identification and correction of mistakes commonly made by interviewers. Piloting should be conducted until all possible problems are addressed and full-scale research can be implemented.

Step 3: Implementation

The implementation of a research project requires the same rigorous attention as its conceptualization, design and planning. Safeguards must be in place to ensure the quality of the data collected, including adherence to guidelines for the selection of respondents, the conducting of interviews and the recording of the responses. Supervisors must be carefully selected and trained on oversight, team management, record keeping, organizing the questionnaire, leading the sampling strategies in the field and double-checking questionnaires. Supervisors must also be able to identify immediately and correct any mistakes made by interviewers in the field while at the same time taking note of the events surrounding the data collection and keeping records of any problems encountered in the field. In practice, of course, many glitches appear during data collection. If there has been adequate preparation, however, few of these will threaten the overall validity of the survey. For instance, in Uganda, some camps initially selected for sampling had to be replaced because of security concerns. This type of constraint is unavoidable and needs to be acknowledged when the final results are presented.

For large population surveys, new technologies can reduce data-collection time and human errors. These technologies have the potential of making data collec-

²⁹ See also, Tim Allen, *War and Justice in Northern Uganda: An Assessment of the International Criminal Courts Intervention* (Crisis States Research Centre, London School of Economics, 2005).

tion more efficient and accurate. Data can be collected using computers or personal digital assistants (PDAs).³⁰ Because data are directly encoded in a digital database, the lengthy process of data entry³¹ is avoided and data-entry errors are minimized. Additionally, new technologies can be used to improve data-collection techniques and analyze new information. PDAs may be combined with global positioning devices (GPSs) to help with mapping and sampling. Use of GPSs can help overcome the challenges of sampling in remote places where no maps, addresses or lists of households exist. Geographic information can also be analyzed using specialized software to detect geographic patterns or more simply for visualization (mapping). For example, a geographic map of victim groups may reveal a pattern of atrocities. Limited access and knowledge and other technical problems currently limit the use of such devices, particularly in (post-) conflict settings where resources are scarce. However, constant technological progress continues to overcome those constraints.³²

Step 4: Data Analysis and Interpretation

Once the data are in a digital format, the database must be examined for any possible error before analysis begins. Data analysis is designed to achieve the objectives and test the hypotheses developed in the study conceptualization. Many problems may appear during data collection and analysis. This being the case, the notes taken by the supervisors in the field are essential for determining the quality of the data collected. Furthermore, an analysis plan for the data should be developed before any analysis takes place. In general, there are two major types of analysis that can be made – descriptive and inferential. Descriptive statistics provide basic graphs, frequency tables, mean, mode (most frequent response), median (the middle score in a list of frequency) and measures of variability.³³ Inference statistics use statistical distribution and tests to examine possible relationships between variables; for example, inference statistics can test the association between type of victims and attitudes toward trials. The appropriate statistical method will depend on the type of measurement,³⁴ the hypothesis being tested and the number of variables involved.

³⁰ Ramesh Krishnamurthy et al., 'Application of Pre-programmed PDA Equipped with Global Positioning System to Conduct Paperless Household Surveys in rural Mozambique' (paper presented to the American Association of Health Informatics, Washington D.C., 2006).

³¹ For data entry of data collected on written forms, the double-entry method is commonly used, whereby the same data are entered by two separate individuals or teams; these two separate databases are then compared to unearth any discrepancies. This process reduces the risk of mistakes occurring when the data are entered.

³² We acknowledge that the use of such technologies is limited in some post-conflict settings. However, the opportunities do exist and researchers now have the ability to 'create' information that was inaccessible before tools such as GPS. The authors are currently in the process of implementing an initial survey using PDAs and are working in such a way as to ensure that the project includes a capacity building component for local researchers. Our work on databases in Uganda has also allowed us to provide a unique mapping of abductions and, for the first time, provide an accurate mapping of the geographic extent of the conflict.

³³ For example, variance, standard deviation, range and quartiles.

³⁴ For example, data can be categorical or continuous. The term 'categorical' refers to data with discrete values (e.g., 'yes,' or 'no'); 'continuous' refers to measurable continuous data (e.g., size).

Interpreting the results is the next step in the study lifecycle. Interpretation often benefits from interdisciplinary team work and discussion with local partners. Researchers must now ask themselves: 'What are the data telling us?' In answering this question, it is important to keep clear the distinction between data interpretation and advocacy. Advocates tend to over-represent evidence that supports their point of view – a pitfall to which qualitative studies may be particularly vulnerable. To avoid this problem, data interpretation must be as neutral as possible. The above-cited Uganda survey suggested that people in the district of Lira were more knowledgeable and supportive of the ICC than participants from other districts. One local partner, who has a long history of working with these communities, suggested that these results could be influenced by the fact that the ICC Prosecutor had specifically announced an investigation into a recent attack in the internally displaced persons (IDP) camps in Lira and that the local district authority had conducted a workshop on the work of the ICC in Lira only three weeks prior to our research. This contextual information may not statistically explain the results, but it offers insights in interpreting the findings of the study. In this case, data interpretation also benefited greatly from the interdisciplinary team of international development, human rights and transitional justice experts from the Human Rights Center, University of California, Berkeley and the International Center for Transitional Justice.

Step 5: Dissemination and Follow-up

In an evidenced-based approach, data are only as good as they are useful. Data collection is a consultation process that fosters participation. Similarly, dissemination and follow-up should be viewed as an exchange, a participatory process. Public presentations should be organized to share the research findings locally with the community and to receive their feedback and interpretations of the data. One-on-one meetings with key decision makers may also be arranged. These meetings provide a platform to share results and discuss the recommendations made, particularly with respect to accuracy and feasibility. However, data dissemination in an area that has been devastated by mass violence and atrocities poses several challenges. Who do we want to hear the results and what is the best way to disseminate the findings? The report may need to be written in local languages. Access to online resources, where it exists, may be limited and the medium used to diffuse the report must be adapted to the local capacity (e.g., a public report versus a peer-reviewed journal). New technologies such as email – or CD-Rom or digital libraries for those without email – may prove useful to ensure rapid diffusion. Regardless of the dissemination vehicle, to achieve maximum impact, reports should: (i) focus on a few messages and in some instances only one; (ii) be grounded with objective data, interpretation and references; and (iii) target a specific audience. It is often useful and even advisable to arrange several meetings with the participants as the format of the research may be too technical. They should have the opportunity to think about the data and engage in a dialogue with the researchers. Too often, the target audience is bureaucrats or

policy makers, which means that those who participated in the study, and who were most affected by the conflict, are excluded from the process that leads to decisions that directly impact their lives.

Validity and Reliability

The last concept we will review in this discussion of monitoring and evaluation research methods are the criteria against which the quality of a study can be judged. Specifically, the two key criteria are the validity and reliability of the research findings. Validity refers to how well a survey or measurement instrument measures what it set out to measure and how well the results can be generalized to a larger population. Reliability assesses whether it is measured consistently. The concepts of validity and reliability are applicable to all types of research and for both qualitative and quantitative studies. In qualitative studies, these concepts are often referred to using different terms, such as ‘credibility’, ‘transferability’, ‘dependability’ or ‘confirmability.’³⁵ While the names may differ, the general idea is the same: to evaluate the quality of a study. There are a variety of different methods that help to ensure the reliability and validity of a study. Good conceptualization and planning of the study, including reviewing questionnaires with experts, carefully planning the selection of the respondents (e.g., sampling), piloting the research and training the interviewers are some of the methods that may be employed to ensure reliability and validity. One of the advantages of conducting quantitative studies is that the validity and reliability may be tested through various statistical methods. Another approach to ensure valid and reliable results is to use concurrently different data collection approaches (e.g., qualitative and quantitative) to determine consistency and accuracy of the findings. Nevertheless, all research has some limitations and it is essential for researchers to explicitly acknowledge them.

Conclusions

This paper builds upon many concepts widely used in program development, monitoring and evaluation outside the field of transitional justice. It presents a framework for evidence-based transitional justice by applying the principles and methods of monitoring and evaluation to the conceptualization, development, implementation and evaluation of transitional justice mechanisms. We illustrate this framework with selected examples of the challenges of collecting information at a population level on the needs for and attitudes towards complex concepts such as peace and justice. Researchers can use monitoring and evaluation research methods to examine such questions as the relationship between population characteristics and attitudes towards peace and justice (e.g., are those most exposed to violence less likely to support trials?).

³⁵ See, Miller C. Delbert and Neil J. Salkind, *Handbook of Research Design & Social Measurement* (Thousand Oaks, CA: Sage Publications, 2003).

Evidence-based transitional justice must rely on valid and reliable information. The framework we present, from research types and questions to a five-step study-lifecycle approach, is designed to provide such information. Although every step is essential to ensure the quality of a study, we emphasize the importance of the conceptualization and planning phase; we also emphasize the importance of developing a thorough knowledge of the issues under study. Such knowledge can be gained through simple tools such as the 'one thousand questions' principle and through the involvement, at every stage, of experts in the field of study and experts in the cultural and social context of the study. Rather than debating the advantages and disadvantages of various research approaches, we argue that each approach contributes to different objectives (e.g., formative evaluation, process monitoring and evaluation or outcome evaluation). We further note that qualitative and quantitative data collection approaches can be complementary rather than competing techniques.

The use of monitoring and evaluation research methods in transitional justice is still in its early stages. However, there is a great need to increase its use given the multiplication of programs and the lack of objective assessments of transitional justice mechanisms that have been developed or will be developed. Participation by local communities and consultation with stakeholders are essential to implement a successful program. Social learning research shows that when individuals feel that they have more control over an outcome, they are more likely to support the process.³⁶ Transitional justice may contribute to social reconstruction if the needs of the population are taken into account in planning, implementing and evaluating mechanisms. Further, researchers and practitioners can use data to understand better how societies rebuild and how or indeed whether transitional justice mechanisms can contribute to this process. Such knowledge is essential to determine if improvements should be made to ongoing processes and to inform the design and implementation of future transitional justice interventions.

³⁶ Albert Bandura, 'Self-efficacy: Toward a unifying theory of behavioral change,' *Psychological Review* 84(1977): 191–215.