Introduction

Although the wounds had healed years ago, you could still visibly see the machete marks indented in his skull, on his forehead and across his ear. This man and the other community members I spoke with, all Tutsi survivors of the Rwandan genocide, had taken part in an internationally funded peacebuilding and reconciliation project I had been hired to externally evaluate. They lived in an urban part of Kigali where other international programs assisted and supported peacebuilding and human rights efforts at a local level, and they had participated in this project that brought together micro-lending, dialogue and psycho-social support for seven years. In this particular meeting, I sat with some of the project participants to discuss their lives in the community and the project's impact on their recovery from the Rwandan genocide more than twenty years earlier. Yet, when I asked this man about the impact the programs had had on him specifically he was not very optimistic: "I do not have the strength to take part, so I do not participate in any savings and loans programs," he told me. "For me I do not see anything worthwhile to invest in my community that can be productive." But, when I turned to a woman in the group, she indicated that things indeed had improved for her after participating in the program. "Before this program I had a house of nine iron sheets, now I have a house of thirty iron sheets," she told me. Their responses, along with those of many others in their group, were often contradictory and confusing to me. It was difficult to ascertain how impactful the programming had been.[1]

It was my job to determine whether these internationally funded programs had helped participating Rwandans improve their situations after they had been ravaged by genocide, war and misery more than twenty years ago.[2] How could I judge, and

[1] All of these interviews were conducted by the author in Kigali and the Southern and Western provinces of Rwanda in December 2015.

[2] For more on the role humanitarian organizations played in the Rwandan genocide, see Uvin 1998 and Rieff 2003.

according to what standards? It was clear that the tools available were not sufficient to make concrete assessments about peoples' experiences. I was given neither the time, nor the close contact with the communities to be able to make confident judgments about their situations. The only standards I could apply to help guide me were developed by outsiders and not sufficient to parse the local context and nuance necessary to understand the impact of programming in communities dealing with very unique problems.

This book addresses the challenges that I wrestled with while conducting evaluations in Rwanda and elsewhere. It makes the argument that beneficiaries are best placed to not only determine the effectiveness of external interventions – policies, programs, and projects – designed to benefit them, but also should be included in the design of the measurement tools used to evaluate them. It also proposes a new and innovative methodology based on inductive, everyday indicators that bridges the divide between quantitative and qualitative approaches to measurement. Using people's own indicators of peace in communities affected by violence is an innovative alternative to existing measurement systems and addresses several of the unanswered questions and criticisms posed by scholars on how the international community can more effectively support localities emerging from conflict to work toward peace. It recognizes the difficulty and often inherent contradictions presented by qualitative assessments done in short timeframes and based on interviews and focus groups, as well as the limitations of the more rigid, quantitative approaches to measurement that attempt to capture complex concepts through simplified quantifiable measures. The study uses mixed methods and participatory frameworks to generate data with the complexity and depth of qualitative findings and the replicability and clarity of quantitative research.

The findings presented here suggest that communities saturated with external interventions after war do not have substantively higher levels of peacefulness than those with lower levels of interventions, according to community-defined indicators of peace. Through the analysis of everyday indicators of peace generated by communities, the study concludes that everyday peace is multidimensional, varies across contexts, and is dynamic and evolving, which is why interventions that address peacebuilding concerns must also share the same characteristics. Therefore, I argue that more intervention is not necessarily better and that more attention must be directed toward the constituent parts and distribution of interventions, and their approach, content and quality, rather than an increase in the size and number of projects. In particular, my findings suggest that conflict-affected communities with large amounts of assistance in reconstruction and development require more interventions pertaining to social cohesion and community social relations than those with little to no assistance.

The study also finds that international and local organizations must develop transparency and coordination among themselves in order to work toward more comprehensive peacebuilding. Currently, communication among intervening organizations

is insufficient due to the competitive nature of funding schemes. I concur with other assessments that donors should incentivize organizations to work together in order to increase the likelihood of success. In addition, local authorities and government officials should be consulted in project implementation where appropriate.

Most importantly, however, I argue that the first step toward determining what works best is to actively include communities not only in the evaluation, and monitoring of external interventions, but also in programming design. By analyzing community-generated indicators and comparing them to existing indicators used by international agencies, I suggest that beneficiaries define peacebuilding effectiveness differently than external interveners and that exogenous and indigenous indicators need to be harmonized in order to more effectively design projects and determine peacebuilding effectiveness at the local and national level.

Most indicators upon which measurement systems rely are designed by researchers and policymakers in capital cities of developing countries in the Global North, and purport to measure communities' progress according to standards that are defined and developed by community outsiders.[3] This is not surprising, considering that communities receiving aid are often left in the dark regarding aspects of decision-making from above and have little understanding of the logic and reasoning behind what they are receiving, or the origin of the services and goods.[4] As I discuss in more detail below, studies have shown that local and international perceptions often differ greatly on whether or not interventions have been successful.[5] Studies have also shown that developing countries implementing reforms often suffer from "isomorphic mimicry," the tendency to introduce reforms that enhance external legitimacy and support in order to ensure external financing. Scholars argue that isomorphic mimicry creates capability traps by focusing implementers on external standards rather than internal impact, ultimately resulting in failed programming.[6] These kinds of dangers could be circumvented by using participatory and community-generated impact indicators that are representative of grassroots concerns.

Everyday peace indicators are the signs we look to in our daily lives to determine whether we are more or less at peace. These are indicators everyone, whether from the Global North or South, subconsciously uses and collects everyday. They can be used as a hermeneutic tool of indigenous technical knowledge to measure and analyze daily life. Everyday indicators are usually quite simple and deal with various aspects of our lives depending on the community we live in.[7] Indicators can vary from hearing barking dogs at night, to the coroner removing dead bodies from the street in a timely fashion, to being attended promptly by a doctor when you are sick, or being

[3] Merry et al. 2015b: 17.
[4] Branch 2008; de Waal 1997; Finnström 2008.
[5] Mac Ginty 2011b; Moore 2013; Autesserre 2014; Richmond 2005.
[6] Andrews et al. 2013: 234.
[7] Mac Ginty 2013a.

able to attend a village festival. By using everyday indicators, I demonstrate the utility of finding ways of including affected populations' voices into more global standards and measurement systems, thereby addressing some of the seemingly innocuous, but often insidious, politics behind existing global systems of knowledge production. By allowing people and communities to decide what indicators determine impact and peace, we shift the inherent power imbalance away from international interveners and researchers to allow the beneficiaries of international interventions to make important decisions for themselves about what constitutes peace. By encapsulating local knowledge within indicators, we are still able to communicate effectiveness in a technical language the international community can comprehend.[8]

This chapter is meant to give a brief introduction to the chapters of the book where many of the arguments and issues presented here are elaborated upon in more detail. I start with a summary of the challenges scholars and practitioners confront when defining difficult to measure concepts such as peace and give an overview of my *big-P* and *small-p* peacebuilding distinctions, which attempt to provide some nuance in the efforts to define peacebuilding. I then proceed to discuss the role and importance of indicators in measurement and the conceptual debates and divisions surrounding the issue of indicators. I move on to give an overview of the methodology and research design for the study and analysis on local level peacebuilding effectiveness illustrated in this book. The overview of methodology and research design is only a brief summary of Chapters 3 and 4, which provide much more detail. I then summarize the main arguments presented in the book based on the analysis of the everyday indicators and survey results in the matched case research design discussed in Chapters 5 and 6. I then elaborate on how these results relate to the literature and theory on peacebuilding effectiveness. I continue by explaining why we need local standards and why a reevaluation of our approach to traditional top-down measurement and indicators developed by outsiders is particularly urgent now. I conclude by discussing terminology used in the book and give a summary of each chapter.

I.1. THE DIFFICULTIES OF DEFINING PEACE

While I was in Rwanda, I was also confronted with the challenges involved in measuring and making judgments about encompassing concepts such as peace and reconciliation. When I asked groups of survivors in Rwanda about their relationships with neighbors and with those that had killed their family members, the responses were varied. Initially, people would respond, "No problem, everything here is fine," and usually they would go on to say, "We are reconciled, maybe not 100 percent, but we are okay. There is no genocide ideology here." Yet, once I dug a little deeper,

[8] See Uvin 2013: 51 for more about how locals and the international community prioritize the same things, but are unable to communicate these to each other effectively.

it was clear tensions still existed. For example, a woman in the same survivor group in Kigali explained to me that when she went home to the place where her family was killed, she would interact and greet her neighbors during the day. But, later she admitted that fear continues to pervade her community. "I trust them during the day, but at night I am afraid. You cannot see inside someone's heart."[9]

Scholars and practitioners have long recognized that reconciliation and peace are difficult to measure. The terms may change depending on multiple variables such as context, culture, language, education and history. It is difficult to map the levels of these kinds of basic variables because they change according to context. Experts thus have difficulty arriving at decisive definitions of what exactly the terms mean and what needs to be measured in order to make claims about them.[10] Qualitative researchers are especially concerned with conceptual validity, or capturing the diverse dimensions of a concept, and are often critical of quantitative indicators that can fail to represent all dimensions of a concept.[11] Measurement is difficult because big concepts such as peace and reconciliation are multifaceted and varied by context and they may contain elements, such as feelings and relationships, which are not easily quantifiable.

Such difficulties are especially prevalent in monitoring and evaluation efforts, where implementers are required to establish indicators to assess the impact of their work, but also in more overarching efforts to measure peace such as producing indices and barometers. Indicators are usually determined by literature reviews and an assessment of already established indicators created by experts, instead of by consulting communities directly. Most organizations and measurement systems continue to use externally developed indicators despite the fact that scholars have found that there are significant differences between local and international actor's narratives and definitions of peace.[12]

Admittedly, part of the problem is that there is little consensus on what peacebuilding actually is.[13] Moreover, there is no existing consensus on peacebuilding effectiveness or what constitutes success or failure, although this issue is being actively pursued by researchers and practitioners.[14] By turning to local communities to elicit their own indicators, the innovative approach presented here allows us to obtain a clearer picture of local understandings of the impact of interventions concerned with normative goals of "peace" and "reconciliation," and helps us understand differing conceptions of these mercurial terms.

[9] This was only the beginning of what would be a very difficult evaluation of reconciliation programming in a complex society with rigid ideas about victimhood and historical memory. For more discussion on the complexity of reconciliation in Rwanda, see Thomson 2013; Davenport and Stam 2009.
[10] Diehl and Druckman 2010: 113; Esser and Vanderkamp 2013.
[11] Goertz and Mahoney 2012: 130.
[12] Autesserre 2014; Richmond 2005; Mac Ginty and Firchow 2016.
[13] Barnett et al. 2007.
[14] Diehl and Druckman 2010: 93–133.

This study distinguishes between *big-P* Peacebuilding and *small-p* peacebuilding. *Big-P* Peacebuilding encompasses all community-level interventions, from humanitarian assistance received immediately after war to longer-term assistance in economic development, health and education, governance reform, conflict resolution, rule of law, transitional justice and security – essentially, everything that purports to work toward a normative goal of peace. In contrast to *big-P* Peacebuilding, the *small-p* approach to peacebuilding is one that is focused, often at a more local level, on agency and the transformation or building of relationships with normative goals of peace.[15]

The book addresses a dilemma raised by scholars and practitioners alike over the challenge of harmonizing "the local" and "the international." It demonstrates that there are ways in which peace can be measured at local levels that are globally transferable. It also investigates the ways in which the international human rights and peacebuilding communities grapple with issues of measurement, monitoring, and data collection and information sharing, and explores the consequences of current monitoring and evaluation practices for the beneficiaries of peacebuilding and development programs. It is fundamentally an academic inquiry into peacebuilding effectiveness at the local level, but deals intricately with methods and methodology since it presents a new approach to establishing peacebuilding effectiveness. In sum, it critically examines questions of power and agency in the threading together of local, state and international needs.

1.2. THE POWER OF INDICATORS

Indicators are powerful tools that wield significant authority in international politics.[16] Global indicators are often designed to name and shame by boycotting and benchmarking states that rank low on indices that measure everything from corruption to human trafficking, or by relegating projects and institutions as delinquent if they do not measure up. They can also keep states, institutions or programs in check by monitoring their progress and behaviors over time.[17] Indicators are a simplified form of a concept and are generated in order to move from a concept to concrete data that can be quantified.[18] Merry et al. (2015b) define indicators as

> a named collection of rank-ordered data that purports to represent the past or projected performance of different units. The data are generated through a process that simplifies raw data about a complex social phenomenon. The data, in this simplified and processed form, are capable of being used to compare particular units of analysis (such as countries or institutions or corporations), synchronically or over time, and to evaluate their performance by reference to one or more standards.[19]

[15] For more of my discussion on this distinction, see Chapter 1.
[16] Kelley and Simmons 2015; Merry et al. 2015; Broome and Quirk 2015.
[17] Kelley and Simmons 2015: 68.
[18] Goertz and Mahoney 2012.
[19] Merry et al. 2015b: 4.

Indicators wield significant power and they are of crucial importance for the procurement of resources, funding, prestige and livelihoods. Indicators can be used to gather qualitative data, but are mostly used in generating quantitative measurement tools. Qualitative indicators typically deal with feelings or intangible issues that cannot be counted.

Gary Goertz and James Mahoney explain the tensions in concepts and measurement efforts between qualitative and quantitative approaches.[20] Qualitative approaches typically resist over-simplification and are concerned with conceptual validity. The failure of indicators to represent all dimensions of a concept is problematic for qualitative researchers because they are concerned with all attributes of a concept.[21] As Goertz and Mahoney put it clearly: "They [qualitative researchers] believe that concepts must be defined independently of data considerations. The definition of a concept should not be driven by the data that are available to measure that concept."[22] Quantitative approaches, however, are driven by data considerations as well as conceptual validity, therefore indicators can be discarded if there is no data available to measure them. Quantitative research is less concerned with measurement error to do with the definition and structure of concepts and instead more concerned with the operationalization and use of indicators, because measurement must be uniform across cases.[23] Although quantitative researchers are concerned with measurement equivalence, or the comparability of measured attributes across populations, obtaining truly uniform data across populations is rare and extremely challenging.[24] Therefore, measurement error for quantitative research occurs at the level of indicators and not at the level of concepts (as it does for qualitative researchers). In fact, some quantitative researchers go as far as claiming that a concept is defined by the indicators used to measure it.[25]

Since indicators are typically data-driven, they can obscure and erase certain elements of the concepts that are being measured. Anthropologists Sally Engle Merry and Susan Wood demonstrate an example using children's rights indicators in Tanzania.[26] They outline a "paradox of measurement," in which they illustrate the tensions between qualitative and quantitative research. Merry and Wood take issue with the fact that what is being measured by indicators is "already recognized as measureable," while other elements of a concept that may be obscured are therefore immeasurable.[27] They give the example of the "right to play" indicator used by the UN Committee on the Rights of the Child and demonstrate how this indicator cannot be measured the same way in industrial contexts as it is in developing contexts like Tanzania since what constitutes "children playing" for us may not apply

[20] Goertz and Mahoney 2012; Mahoney and Goertz 2006.
[21] Goertz and Mahoney 2012: 130.
[22] Ibid.
[23] Mahoney and Goertz 2006: 244; Sartori 1970.
[24] Davidov et al. 2014: 55–61.
[25] Ibid.
[26] Merry and Wood 2015: 206.
[27] Merry and Wood 2015: 207.

the same way in other places. Merry and Wood question whether this indicator adequately captures the reality of families with children in Tanzania and whether there is a common understanding of play across cultural contexts. They conclude that in order to generate data comparable at the international level, researchers were forced to forge ahead with indicators that were not culturally appropriate.[28] In other words, top-down approaches to designing measurement tools only include indicators that are accessible or exposed to experts and researchers because they have already been discovered by the literature or other studies. Merry and Wood demonstrate, through their case study, that the use of indicators requires interpretation or "translation" and these translations pose challenges to measuring phenomena not previously counted. Therefore, issues are obscured, and hence immeasurable, since it is impossible to measure what you don't know is there. Their frustration with the data-driven approach is exactly what Goertz and Mahoney identify as the tension between quantitative and qualitative approaches in concepts and measurement.

The paradox of measurement is a problem particularly for those attempting to measure at a local level using indicators developed by outsiders, because these indicators may only represent certain issues within the community that are either already being measured or are typically measured to ascertain the effectiveness of programming (e.g. morbidity or infant mortality). This paradox is also illustrated by the difficulties presented to researchers in accurately reflecting gendered interests and problems. Gendered indicators – such as life expectancy, average wages and access to resources – often neglect issues of specific importance to women (or to men) because they do not take into account their specific circumstances or the relevance and meaning of these existing indicators to their wellbeing.[29] In addition, such top-down approaches often only include indicators that are easily measurable or already measured, discarding indicators without existing data or where data collection is too difficult.

The paradox of measurement is also fundamentally a clash between positivist and interpretivist approaches. Fred Schaffer explains how positivist social scientists find ways to reconstruct everyday words to meet their research needs, which require measurement, comparison and generalization.[30] He shows that positivists strip everyday language from its sometimes vague, multidimensional meanings in order to generate a specialized, technical language that allows for tangible results and analysis. For example, the term democracy may have a very fixed and engineered meaning for positivist social scientists that could mean something entirely different for people experiencing democracy in practice or in a foreign context. Andre Broome and Joel Quirk similarly demonstrate that this is done in global benchmarking efforts through what they call reification or the

[28] Merry and Wood 2015: 214–15.
[29] Austen, Jefferson and Thein 2003: 2.
[30] Schaffer 2016: 5.

> translation of complex phenomena into observable and quantifiable conceptual categories that are presumed to be universally applicable irrespective of cultural or historical context. Reification effectively stabilizes the meaning of complex and highly contested categories, such as democracy, freedom and stability.[31]

In other words, in order to measure a concept, social scientists need to be able to pin down exactly what it is and which elements belong in the conceptual data container to be measured.[32] This involves making important decisions about what belongs and what does not, as well as generalizing significantly in order to make the indicators universally applicable.

Regardless of the empirical approach, concepts of interest need to be defined and a method for making systematic observations developed. However, in the case of interpretivist scholarship, the primary conceptual task is to elucidate shared meanings and interrogate the relationship between social science language and everyday meanings,[33] or the distinction made by Clifford Geertz between experience-distant and experience-near concepts.[34] Geertz distinguished between these two concepts in order to demonstrate how social science creates abstractions in order to study phenomena from a distance. Experience-near concepts encompass the ways in which those directly experiencing phenomena might describe them. For example, referring to democracy as "the majority of people who voted for something won." Interpretivist social scientists also must construct concepts, but use inductive methods to glean the applicability of a concept to a particular place or culture. Both interpretivist and positivist scholars say things about concepts they have made decisions about and create experience-distant concepts to analyze and communicate the experience-near. Both approaches have their strengths and weaknesses. Because of their interest in good numerical measures, positivists are able to make claims about how large populations view their constructed concepts, whereas interpretivists are able to construct concepts for smaller populations, which makes them more concerned about substantively valid concepts.[35] Scholars using a combination of methods are often the most effective because they are able to give context to generalizability.

It is the dynamic debate between qualitative and quantitative and positivist and interpretivist approaches that I address here and is one of the main areas where this book makes a contribution to the literature.[36] Just as Schaffer calls for "an approach that provides people opportunities to articulate the connections that they themselves

[31] Broome and Quirk 2015: 828.
[32] Sartori 1970: 1052
[33] Schaffer 2016: 10.
[34] Schaffer 2016: 2; Similar to emic and etic understandings of sociocultural phenomena. See more discussion on emic and etic approaches to understanding culture, in Avruch 1998: 60–5.
[35] Goertz 2005: 2.
[36] I am in agreement with Schaffer's caution that these categories are not fixed and inflexible, but are illustrative of the overarching methodological approaches scholars tend to fall into. See footnote 1 in Schaffer 2014 for more on this.

make between meanings, the complexities that they themselves grapple with, and the conceptual puzzles that they themselves have not been able to solve,"[37] I argue that quantitative and positivist approaches can and must take concepts more seriously and find ways to integrate more comprehensive, democratic and participatory elements into measurement and monitoring.

Doing so is crucial precisely because of the political, "soft" power wielded by indicators, especially with the onset of the data revolution and the inordinate dissemination of information globally. The data revolution has made an enormous amount of information available to researchers. With the radical increase in the availability of information, indicators should now work alongside concepts to offer more holistic representations of social phenomena since they are no longer as restricted by data availability. Indicators are the metrics that increasingly matter in a world of new public management and technocracy. A change in the design of indicators is particularly urgent since increasing numbers of policy outcomes are influenced by indicators. Access to indicator-based data allows policy-makers to quickly scan results rather than read underlying reports, which can take weeks.[38] This increased dependency on indicators to guide policy outcomes makes the generation of concept-rich indicators more urgent than ever.

1.3. A FEW WORDS ON RESEARCH DESIGN AND METHODOLOGY

This study employs a matched case research design[39] and uses mixed methods in two very different contexts to investigate whether interventions after war indicates effects on community-generated indicators of peace.[40] The study uses an innovative methodology that allows researchers and others to measure impact according to the ways individuals themselves measure peace in their communities. Instead of drawing on indicators of success developed by "experts" and "scholars" – often in the Global North with data collected in the Global South – researchers ask communities themselves to establish their own indicators of peace, which are then measured longitudinally through surveys. The resulting everyday indicators articulate, translate and vernacularize the measurement of local data in ways that traditional measurement systems cannot. The conclusions of this book are based on survey data from 2,038 surveys in four communities in Uganda and Colombia.[41] The study uses the matched case research design in four communities supplemented by over one-hundred interviews with community members in the villages, local and national elites in

[37] Schaffer 2014: 328.
[38] Kelley and Simmons 2015: 57.
[39] Maclean 2010: 32.
[40] See Chapters 3 and 4 for much more detail on research design and methodology.
[41] Indicators of peace and reconciliation were collected in Colombia. The analysis in this book presented in Chapters 5 and 6 is based solely on the indicators of peace.