

The Complex Task of Advocacy Evaluation: Lessons from Brazil

William N. Faulkner¹

¹Plan, São Paulo, Brazil



Abstract

Much is written about the obtuse application of evaluation tools without appropriate theoretical understanding, but analysis of situations wherein theories remain unapplied for the lack of appropriate tools is rare. The paper introduces the tools brought to bear during two advocacy evaluations, analyzing their practicality, including how they push back on the theoretical demands of advocacy evaluation.

Keywords

Advocacy, Qualitative Analysis, Software, Case Studies, NGOs

Glossary

Abbreviation	Meaning
ANCOP	<i>Associação Nacional de Comitês da Copa</i> - National Association of World Cup Committees
FEC	Forced Evictions Campaign
M&E	Monitoring and Evaluation
NGO	Non-governmental organization
PlanPP	Plan Políticas Públicas

Introduction

This paper summarizes some of the lessons learned during two advocacy evaluations by Plan, both of international campaigns by non-governmental organizations (NGOs). The center of gravity of both evaluations lay in Brazil, but both focused on advocacy campaigns with global dimensions being carried out on multiple continents. Advocacy evaluation has distinctive characteristics, including the usual shape of efforts (networked instead of hierarchical), the patterns of expected change (punctuated equilibrium instead of linear), and the sensitivity of the information involved (most advocacy efforts have dedicated opponents).

The primary goal is to tie the incipient theoretical literature on advocacy evaluation to the concrete experience of implementing an evaluation methodology. While much is written about the inappropriate application of tools by evaluators not properly versed in the theoretical background of a particular advocacy field, less attention seems to have been paid to the other side of the coin, wherein theories remain unapplied for the lack of appropriate tools. Following the ideas of Latour (1987, 2004), evaluation tools and technology should not be regarded as passive instruments to be dominated, but instead as active ingredients which shape and underpin theory.

To cope with the peculiarities of the task, Plan developed and employed some novel tools, in particular, a way to organize qualitative data with mind mapping software, and a basic combination of tools for visualizing data on connections between strategies and outcomes. This paper details how these tools were formed as well as their advantages, disadvantages and the conclusions which emerged from their application. By covering the steps of the evaluation project from inception through the dissemination of results, this paper seeks to analyze some of the dimensions of their practicality and how they both respond and push back on the theoretical demands of advocacy evaluation.

The first section explains the background of the evaluation projects; the second section covers the key aspects advocacy evaluation as recognized in the incipient literature on the subject; the third section details the two unique tools used to analyze qualitative and quantitative data gathered during both projects: (a) sankey diagrams, and (b) mind-mapping software; the fourth section discusses strategies around dissemination of results produced with these tools, and the final section abstracts some of the more general lessons learned.

Project Background

Instituto Sou da Paz ('I am for peace' Institute)

An NGO, Sou da Paz is a household name in their home base of São Paulo for their violence prevention programs in the city periphery working with youth and police officers, and their central role in a national disarmament campaign. Plan contracted with Sou da Paz in late 2012 to evaluate an international branch of their activities based on several staff members'

involvement in advocacy efforts surrounding two UN processes regulating the trade, storage, and use of conventional weaponry since the early 1990s.¹

WITNESS

WITNESS, an NGO based in Brooklyn, USA, developed and undertook a Global Forced Evictions Campaign (FEC) between June 2011 and May 2014 with funding appropriated by the Adessium Foundation. PlanPP worked with WITNESS between November, 2013, and February, 2014, to accomplish the final evaluation of the FEC, whose primary objective was to “protect the rights of poor and underrepresented communities to housing, livelihood and community from forced evictions by development.” (WITNESS Grant Proposal, p.1) WITNESS provided support to 24 pre-existing local and regional groups in Brazil, Cambodia, Egypt, India, and Mexico in the form of a suite of activities developed and modified across geographies and over time. These activities involved providing equipment, organizing in-person meetings and trainings, directly assisting in the production of videos (i.e. scripting, filming, editing, etc.) and indirect support in the form of manuals and consultations via phone/Skype. The evaluation focused on Mexico and Brazil, using a small amount of data gathered from the other countries to contextualize the findings in these cases.

Relevant Aspects of Advocacy Evaluation

Advocacy evaluation is an “elusive craft,” (Teles and Schmitt, 2011) not only due to the fact that it is a new and rapidly evolving field, but also because the sophisticated methodological toolkit brought to bear on social programs aimed at product or service delivery is widely inapplicable to most strategies aimed at political influence. The incipient literature on the subject (commissioned mainly by a few foundations attempting to develop strategies with which to improve the results of their advocacy-dedicated portfolios) highlights three features relevant to the cases study presented here:

Expected Pattern of Change

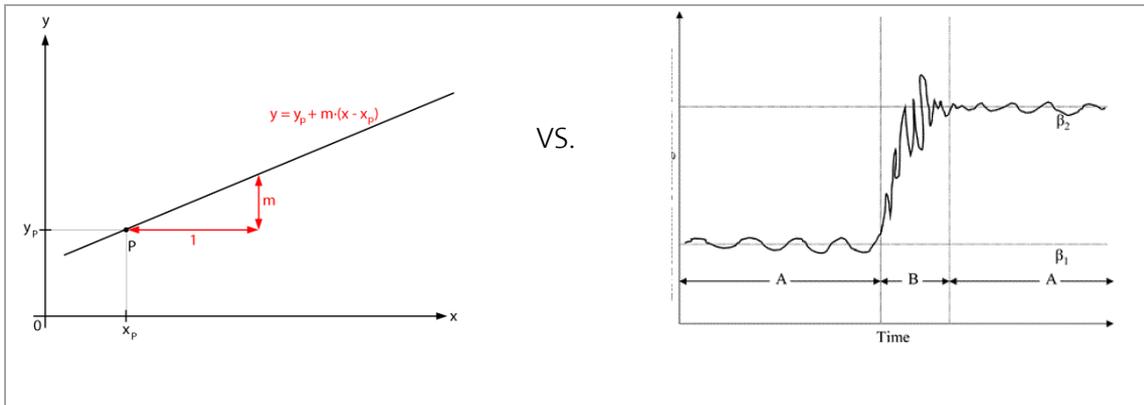
Advocacy efforts are normally characterized by a multiplicity of activities aimed at a broad swath of intermediate outcomes. The pace of advocacy often follows a ‘punctuated equilibrium,’ whereby smooth, linear change is replaced by long periods of stability followed by brief spurts of intense activity (see Figure 1: *Linear Change vs. Punctuated Equilibrium*):

“While policy victories can sometimes seem quick because they occur in sudden windows of opportunity, they often require years of groundwork and several attempts and failures.” (Beer, 2012, p. 2)

Precise timing is therefore even more crucial than in standard program evaluation. Information may only be relevant for a few weeks, days, or even hours. Taking the time to lay a base of empirical information is important, but the evaluator(s) must be open and prepared to accommodate emergent demands, sometimes quite suddenly (Coffman and Beer, 2011, p. 7).

¹ As opposed to nuclear, chemical, or biological weaponry, or cluster munitions, around which international legislation and regulations had already been passed.

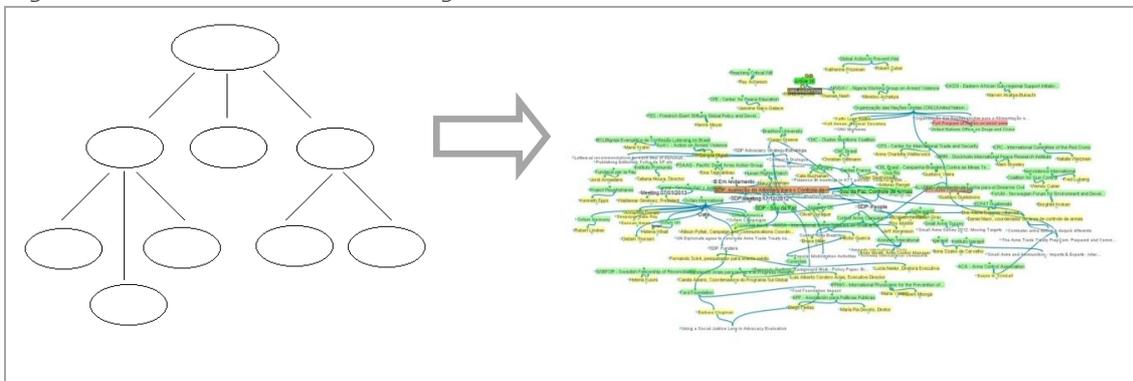
Figure 1: Linear Change vs. Punctuated Equilibrium



Expected Organizational Structure

Advocacy networks tend to eschew centralized, hierarchical structures in favor of diffuse webs, an aspect to which the methodologies for both collecting and disseminating information must adapt. Like all evaluations but perhaps with even greater sensitivity, the appropriacy of methods for advocacy evaluations depends on their spatial/social/cultural context. To the degree feasible, advocacy evaluation results should be possible to disaggregate by strategy, time-period, actor, and other pertinent dimensions so that stakeholders within the network can extract the information most relevant to their particular interests (Weiss, 2007).

Figure 2: Hierarchical vs. networked organization ²



Sensitivity of Information

As highlighted by Patton (2008), the disclosure of results of an independent advocacy evaluation, particularly when occurring in an ongoing context, is sensitive affair. The 'independent' nature requires that the evaluator maintain strict separation between the identity of the respondents and the information they provide. The 'advocacy' aspect, however,

² The network visualized here was constructed with empirical data on Sou da Paz's international advocacy network with the same PersonalBrain software used to organize the interview citations (discussed in

stems broader concerns: many advocacy efforts have serious, dedicated opponents. Disseminating results must consider the possibility of harming the evaluand by revealing strengths and weaknesses to the opposition. As a demonstration of the effect of a sensitive, multi-actor advocacy space, Sou da Paz asked that a previous version of this paper prepared for the Rede Brasileira de Monitoramento e Avaliação (Brazilian Monitoring and Evaluation Network) in October, 2013, use a pseudonym for the organization and avoid references to an specific organizations or groups besides the UN. WITNESS, on the other hand, showed no such concerns; as the first campaign to raise awareness of forced evictions at a global level, the FEC unfolded in a relatively uncontested advocacy space with a few, disparate partners and no dedicated opposition lobby.

Concretely, the previous two aspects implicate that a single dissemination product will likely not suffice for all of the intended audiences. The evaluation team may have to produce different versions of the data collection instrument(s) and/or results ‘packages,’ tailored to the specific needs and interests of stakeholder sub-groups.

Expected Outcomes and Evaluation Conclusions

The final outcome of advocacy, often the implementation and enforcement of a regulation, law, or treaty, can take years to manifest, if indeed it ever does. Even in the event of success, establishing a causal link between advocacy and the outcome is usually difficult if not impossible. The focus of the evaluation will usually be on intermediate outcomes and process – the small steps and the paths taken between them (Beer, 2012; Coffman, 2003; Forti, 2012).

Evaluation Methodology

Data Collection

Two key factors enabled the PlanPP team to couple the subtle, complex demands of these projects with the practical restraints on the time and budget. First, early on the PlanPP team jointly established appropriate expectations for the project outcomes with the receptive and flexible leadership at Sou da Paz, WITNESS, and the Adessium Foundation. As such, all parties clearly understood from the outset that no impacts or causal relationships would be empirically measured. Instead, the goal would be to provide a descriptive assessment of the client’s contributions to intermediate achievements.

Second, both Sou da Paz and WITNESS staff enthusiastically participated in the design of the data collection instruments. The data collection itself was fairly conventional, employing an email questionnaire and semi-structured interviews. Tailored versions of the survey and interview script were designed for subsets of respondents depending on the sector and preferred contact language of the respondent. Semi-structured interviews were conducted to deepen and detail the broad-but-shallow information captured by the questionnaire.³

³ Scheduling interviews was greatly facilitated by being located in São Paulo, which, in December and January, lies centrally between European and North American time zones.

Data Analysis

The process of data analysis took on a less conventional form. Much has been written in the field of evaluation lamenting the haphazard and premature application of novel technological tools without the proper grounding in theory – social, mathematical, or otherwise. Perhaps because evaluation is by default an applied field, however, few authors exhaust much effort considering the other side of the coin. Novel theories and fields of evaluation often require tools to reach their full potential in application or sometimes to be applied at all. To construct an analysis based on the still embryonic theory briefly outlined in the previous section, the PlanPP team made use of several crucial tools, only two of which will be discussed here:

Sankey Diagrams:

Aware that advocacy is a multifaceted activity composed of various interlocking components, PlanPP approached the evaluation understanding that asking about or presenting results around universally ‘effective’ strategies would evoke a distorted picture of these efforts. Instead, the evaluation sought a way of displaying information which could exhibit which combinations of strategies were most effective at contributing to specific outcomes. Several questions on the survey questionnaire therefore aimed to elicit responses amenable to visualization in Sankey diagrams, which seemed the most viable and intuitive format to display these types of connections.

Sankeys are a special type of [flow diagram](#). In the format utilized for these evaluation, each column is simply a [stacked bar graph](#): the full height of each column represents 100% of responses and the relative sizes of each section represent the percentage of respondents that chose each option. The relative sizes of each section represent the percentage of respondents that chose each option. Sankeys overlay this information with connecting bars, the widths of which correspond to the percentage of respondents who paired the two connected options in different columns (see Figure 3 and Figure 4).

These figures’ purpose was twofold: (a) to guide probing questions during interviews, and (b) as data sources to be triangulated with interview data to highlight areas of agreement and disagreement. There is little that can be concluded from these diagrams alone, however.

Procedure

On the most practical level, the production of these diagrams is a multi-step, multi-software process requiring significant computer literacy, but requires only one paid software package which most evaluators will already have: Microsoft Excel. The procedure is essentially an exercise in database preparation which transforms survey answers into a format which can be processed by the cDataSet excel template.⁴ The steps vary depending on how and with which platform the survey answers were collected, as well as the user’s level of comfort with Excel formulas.⁵ With basic knowledge of the html language, the visualizations output can also be customized. For example, note that Figure 3 represents a visualization with no alterations to the html code which defines the colors, opacity, fonts, and other aspects of the diagram, while

⁴ Available at the Excel Liberation website: <http://ramblings.mcpher.com/Home/excelquirks/downlable-items>

⁵ For further details on the options for this data processing, along with links to the necessary software packages and add-ons, please contact the author at wnfaulkner@gmail.com.

in Figure 4 small adjustments caused thicker bands to be darker and more opaque, and a larger font size.⁶

Figure 3: Sankey Example – Sou da Paz

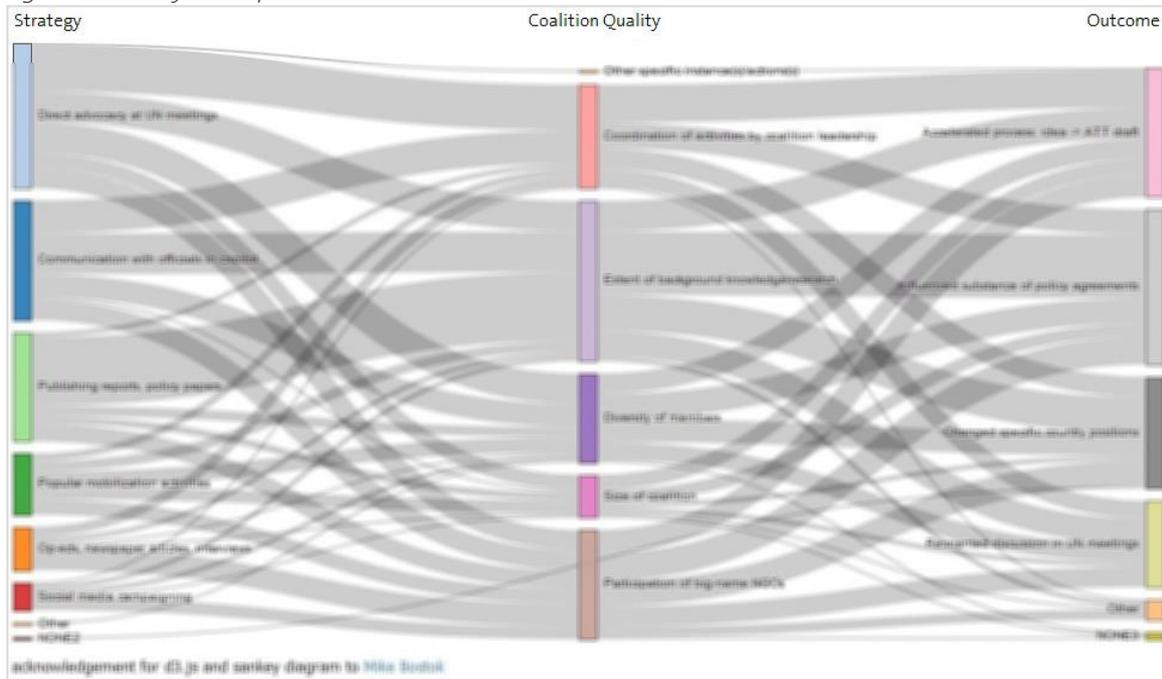
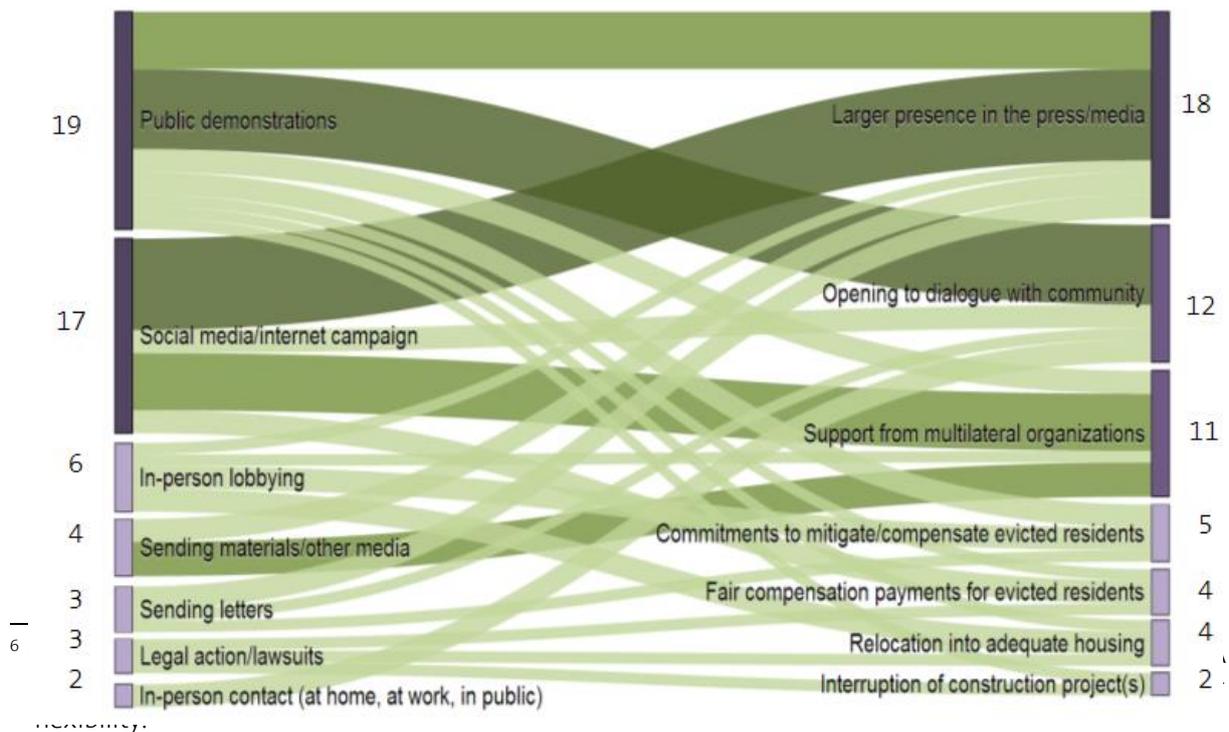


Figure 4: Sankey Example - WITNESS



Evolution

Beyond adjustments to the formatting of the visualization, comments and feedback on the original survey from respondents spurred several substantive changes between the two projects. First, the framing of the survey question (see Table 1) mattered enormously. Simplest and perhaps most practical, the WITNESS survey inserted a standalone page of directions with an example before the Sankey question. The question itself in the Sou da Paz survey concentrated respondents' attention on the overall influence of civil society on UN processes. After taking the survey, several respondents commented that the broad frame made the question difficult to answer, forcing them to consider the average effect of multiple diverse organizations over more than a decade. As reframed for WITNESS, the question asked respondents to focus on their own organization, and this information was then crossed with other information (location, main actions as cited by interviewees, respondent position in the network) to form place and case-specific interpretations.

Example Interpretation of Figure 4:

19 respondents (35%) from Brazil cited public demonstrations or social media campaigning as an effective strategy. These two strategies led to a wide range of outcomes, but were primarily useful in garnering a larger presence in the media and opening officials to dialogue with the community (the two darkest/thickest bands). This evidence supports conclusions the conclusion that ANCOP and its derivatives sit firmly in the 'grassroots' camp, mainly employing public demonstrations and internet campaigning. These partners are not coherent, sophisticated advocacy groups, but loose networks whose access to political fora is mainly mediated by how much "racket"

Table 1: Survey questions designed to elicit information for Sankey diagrams

Sou da Paz (2012)	WITNESS (2013)
<p>What worked? Which combinations of Civil Society's strategies and qualities influenced the UN processes MOST? Below please choose one option from each category to form what you judge to be the three MOST effective combinations:</p> <p>Example: Strategy + Quality ---[influenced]---> Change</p>	<p>How do you think you or your network/movement/organization has been effective in combatting forced evictions in the past three years? Give 1-3 examples of a good strategy (or combination of strategies) you used and the ways it (they) helped combat forced evictions.</p>

Second, the Sou da Paz survey attempted to go beyond strategies to understand the necessary and sufficient factors for their effectiveness. The survey asked respondents to highlight combinations of "strategies" and "civil society's qualities" which led to specific changes. Unfortunately, many respondents dubbed the question both confusing and overcomplicated, and many left it unanswered as a consequence. To simplify, the WITNESS survey forwent the additional information regarding the conditions of a strategy's effectiveness, and requested only that respondents pair a strategy with an outcome.

Third, the Sou da Paz survey asked a separate question (in the same, Sankey-friendly format) with unique response options about Sou da Paz's contribution to the civil society coalition.

Although respondents gave no direct feedback on the combination of the two questions, the evaluation team interpreted the lower response rate to the second question as at least in part due to frustration with the question format. The team also recognized a missed opportunity, in that the unique answer options for each question provided no simple way to connect the data from both questions in a meaningful manner. To mitigate these issues, the WITNESS survey followed up the Sankey question with another asking respondents to rate the contribution of the specific dimensions of WITNESS’ support to the effective strategy-outcome pairings on a Likert scale, permitting a detailed disaggregation of the effectiveness of WITNESS’ involvement.

Mind Mapping:

Another challenge the PlanPP team faced was how to synthesize qualitative information from 30+ (Sou da Paz) and 23 (WITNESS) interviews into useful and pertinent analyses while allowing for novel, unexpected categories and results. PersonalBrain, a mind-mapping software, includes several features which allowed the team to at least partially reconcile ‘top-down’ and ‘bottom-up’ categorization, using the following procedure:

- Transcribe interviews into a word processing software, translated into English when necessary
- Insert semi-colons and copy-paste the text into ‘thoughts’ (references)
- Connect as sibling nodes to the speaker/source
- Tag and color-code according to speaker and/or reference attributes. In Sou da Paz, the thoughts were assigned different colors if their speaker came from government or civil society. In both projects, references which could betray the identity of the speaker were tagged as “confidential” so as to be able to easily remove them and pass the complete, de-identified database to the client.
- Once all interviews have been transcribed, go through node-by-node and attach each as a child to relevant (sub)categories. First, basic categories were created according to conversations with the client and based on the terms of reference. Second, subcategories were added in nested levels and thoughts assigned until each subcategory contained relatively few (generally < 20), easily analyzed references.

Figure 5: Visualization of Sou da Paz Quotations Database



Advantages and Disadvantages

With respect to other methods and tools for qualitative analysis, this particular approach holds some unique advantages, and some all-too-familiar disadvantages, summarized below.

Advantages

- PersonalBrain excels at handling and visualizing infinite-level nested hierarchies, allowing the user to quickly expand/collapse different levels. To mimic the function of coding a single piece of text to multiple themes (i.e. ‘nodes’ in NVivo), one can attach a reference to various parent nodes.
- Repeatedly reading and manually categorizing the citations enforces a deep familiarity between the evaluator and the dataset.
- The end product, de-identified and culled of sensitive or identifiable quotes, is a valuable product to supply the client alongside a report, one which allows them to access and further explore the raw qualitative data themselves..
- To be able to view the references in the cascading hierarchical format requires PersonalBrain Pro (one-time license US\$219). Although not free, it is still much cheaper than most licenses for qualitative analysis packages (less than one-tenth of the fee for a new commercial license for NVivo).
- Although there is a learning curve as with any software, the interface for PersonalBrain is much more tactile, generally based on drag-and-drop commands, than many menu- or code-based qualitative packages.

Disadvantages

- Undertaking the above qualitative analysis is hugely labor intensive, requiring multiple times the work hours compared to highlight coding-based softwares like NVivo for a given amount of data.
- Thoughts can contain only 256 characters, although further text can be copied into the "Notes" field of the reference that will show up with a mouseover.

Results Presentation

Report Structuring

The advocacy networks surrounding Sou da Paz demonstrated the characteristic punctuated patterns of change, networked organization, and consciously compartmentalized sharing of information discussed previously. This context powerfully influenced the design of the evaluations products and their dissemination plan. In WITNESS’ case, three factors diluted the influence of the ‘advocacy’ aspects on the product design: (a) detailed pre-planning had occurred before PlanPP’s contracting, and adherence to this plan was clearly expected, if not strictly enforced; (b) the strong desire to marry the analysis to the evaluation questions; (c) the dearth of an obvious consolidated adversary to the FEC (and hence many of the sensitivity concerns which surrounded the Sou da Paz project).

Figure 6: Example of Claim-Evidence Result (Sou da Paz)



The evaluation team chose to structure the final report for Sou da Paz by inserting [claim → evidence] pairings into a structure with modules directed at specific audiences (example below). The reasoning behind this structure was:

- In such large, complex environments there could be no ‘main finding’ of the evaluation, but instead a network of independent results only loosely linked because the system of interest was simply too big, diverse and rapidly evolving to classify or analyze as a whole.
- This module-based structure has the advantage of being easily reconfigured into various versions targeted at different audiences (an example in the case of Sou da Paz being civil society actors versus politicians and diplomats).

Nevertheless, the module architecture concedes the value of a coherent, ordered story, complete with sequenced claims. The free-order of the sections means that claims in separate modules must be standalone, and cannot build on one another. Similarly, the nested claims can only awkwardly be inserted, flattening the structure of the document and impeding the summaries from weaving together nuanced, overall theses.

Discussion & Conclusions

The preceding sections discussed the context, instruments, and presentation for two evaluation case studies, attempting to note the facets which might help readers in deciding

whether or not to apply them in the future. This section extrapolates some of the more abstract lessons, pertinent mainly because advocacy evaluation is still a new field, “stammer[ing], making its debut, creating itself from nothing in direct confrontation with the world,” unlike some older areas of evaluation which already rest “on the indefinite sedimentation of other disciplines, instruments, languages, and practices.” (Latour, 2004, p. 30) Imagining a matrix of methods and tools applicable to advocacy evaluation crossed with all the possible contexts in which advocacy evaluation might occur, only a tiny percentage of the cells would currently be populated with concrete case studies for reference. In this relatively unexplored landscape, both the possibility of discovery and the risks of disaster grow. What conditions potentiate the former and mitigate the latter?

Both project case studies discussed in this paper represent close to ideal circumstances for fostering innovation. First, on the scale of clients served my Plan, both Sou da Paz and WITNESS represented relatively short-duration, low-budget projects. Plan also enjoys a relatively isolated status as Brazil’s largest wholly-M&E dedicated private-sector consultancy (with only 6 full-time staff). The national M&E sector is nascent, devoid of cutthroat competition from comparable entities in a market where demand for reasonable-quality M&E services outstrips their supply. This low-pressure environment curtails worrying about the financial and reputational risks of experimentation. Internally, as small projects within the portfolio, Plan assigned three-person teams in each case, concentrating the majority of decisions and implementation and reducing the risk of internal disagreement.

Second, staff members of both clients took great personal interest in the projects, spending long hours helping the evaluation team to refine the data collection instruments, engage interviewees, discuss results as they emerged, review documents, and follow up on requests for further information. The continual communication allowed the evaluation team to also iteratively rearticulate the reasoning and risks behind each component of the project, rather than having to squeeze all of the details into a single meeting or presentation. Third, neither client had much experience commissioning or running evaluation projects, and thus had few preconceptions about what a ‘conventional’ project in their case might entail.

Exploring the possibilities of advocacy evaluation was therefore in these cases greatly encouraged by a somewhat protected environment, but one which also tied together individuals with some essential common characteristics. As junior members of the evaluation profession, the evaluators had less to risk in terms of individual reputation. Having few ties and commitments outside of these projects was also essential during some moments in allowing the team to accommodate and respond to emergent demands, and in general to the tight timeline of both projects.

The lack of prior experience in the area also forced the team to relinquish notions about the intended purpose of the tools used to implement their strategies (e.g. PersonalBrain, MicroSoft Excel). Instead, the team focused on their components and functions and then tried to reconcile the potential configurations of these functions with what the theory implied. When these tools lacked small but essential features, the generosity of the web community supplied them for free (e.g. cDataSet). Still, uncovering these add-ons, plugins, and programs required figuring out the key search terms which other fields used to described the desired functions (few if any had been applied in evaluation).

Not all advocacy evaluations will occur under such propitious contexts, nor will they all involve teams with the aforementioned characteristics. As an emerging field, particularly outside of the Global North, however, advocacy evaluation requires rethinking old patterns and creative solutions to the unique challenges which it presents in both theoretical and practical domains, thus capitalizing on these opportunities is essential in moving the field forwards. Plan hopes that recording and sharing these experiences may benefit future advocacy evaluations both within Brazil and globally.

Acknowledgements

The author would like to thank the other members of the evaluation team (and the leaders of Plan), Fabrizio Rigout and Camila Cirillo, for both finding these clients and supporting the development of these rather risky, experimental methods. The Plan team also thanks Mauro Zackiewicz for first suggesting the Sankey as an appropriate data visualization format. Our contacts at Sou da Paz and WITNESS also deserve recognition and thanks for their attentive collaboration on these projects and consent to have information regarding them disseminated as public goods in this paper, so that hopefully other evaluation professionals may learn from their example.

References

- Beer, Tanya. 2012. "Best Practices and Emerging Trends in Advocacy Grantmaking." (January).
- Coffman, Julia. 2003. *Lessons in Evaluation Communications Campaigns: Five Case Studies*. Cambridge, MA.
- Coffman, Julia, and Tanya Beer. 2011. "Evaluations to Support Strategic Learning : Principles and Practices." (June).
- Forti, Matthew. 2012. "Measuring Advocacy – Yes We Can!" http://www.ssireview.org/blog/entry/measuring_advocacy_yes_we_can (February 13, 2013).
- Latour, Bruno. 1999. *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge, MA: Harvard University Press.
- . 2004. "Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern." *Critical Inquiry* 30(2): 225–48. <http://www.jstor.org/stable/10.1086/421123>.
- Patton, Michael Q. 2008. "Advocacy Impact Evaluation." 5(9).
- Teles, Steven, and Mark Schmitt. 2011. "The Elusive Craft of Evaluating Advocacy." *Stanford Social Innovation Review* Summer: 38–43.
- Weiss, Heather. 2007. "Advocacy and Policy Change." *The Evaluation Exchange*.
- WITNESS - Grant Proposal. (2014) (pp. 1–16). Brooklyn.