Pre-Analysis Plans for Qualitative Research

Planes Pre-Análisis para la investigación cualitativa

RAFAEL PIÑEIRO
Universidad Católica del Uruguay

FERNANDO ROSENBLATT
Universidad Diego Portales, Chile

ABSTRACT

The design and registration of Pre-analysis Plans (PAP) represents a significant improvement in social science research transparency. This tool is commonly used in experimental research. In this research note, we suggest extending the use of PAP to qualitative research. In recent decades, researchers have produced several methodological innovations, which have improved the quality of qualitative analysis. New tools also have been developed and researchers have taken important steps to improve data collection and transparency in the analysis of qualitative data. The development of Pre-analysis Plan-Qualitative (PAP-Q) aims to synthesize these advances into a guide for researchers, in order to improve transparency and better specify the role of induction in the construction of causal arguments.

Key words: pre-analysis plan, qualitative research, transparency, induction

RESUMEN

El diseño y el registro de Planes Pre-Análisis (PAP) representa un avance significativo en la transparencia de la investigación en ciencias sociales. Esta herramienta es comúnmente usada en la investigación experimental. En esta nota de investigación proponemos expandir el uso de los PAP a la investigación cualitativa. En las décadas recientes, se han producido una serie de innovaciones metodológicas que han mejorado la calidad de los análisis cualitativos. Estas nuevas herramientas han permitido que los investigadores hayan avanzado en el mejoramiento de los procesos de recolección de datos y la transparencia en el análisis de datos cualitativo. El desarrollo de Planes Pre-Análisis Cualitativos (PAP-Q) busca sintetizar estos avances como guía de investigación para mejorar la transparencia y la especificación del rol de la inducción en la construcción de argumentos causales.

Palabras clave: plan pre-análisis, investigación cualitativa, transparencia, inducción

* We thank Thad Dunning, Jake Bowers, Heather Lanthorn, Juan Pablo Luna and Sergio Toro for their insightful comments on previous versions of this paper. All remaining errors are our own.

** Fernando Rosenblatt thanks CONICYT-Fondecyt #11150151 and acknowledges support from the Chilean Millennium Science Initiative (NS130008).
I. INTRODUCTION

Qualitative research has witnessed significant advances in its methodological sophistication, especially in terms of case selection and in the assessment of descriptive and causal inferences. In recent decades, partly as an unexpected reaction to King et al.’s (1994) classic textbook, various scholars published methodological manuals that have sought to highlight the specificity and virtues of qualitative research. In that context, there has been an impressive development of analytic tools and guides for data collection (Ragin 1987, 2000, 2008; George and Bennett 2005; Goertz 2006; Rihoux and Ragin 2009; Brady and Collier 2010; Gerring 2012; Goertz and Mahoney 2012; Schneider and Wagemann 2012). These developments include the formalization of methods that build upon mathematics and logic. A noted example is the formalization of the process tracing tests (Mahoney 2008; Zaks 2011; Bennett and Checkel 2015). More generally, the qualitative tradition has highlighted the analytical value of mechanistic causation, already developed by Elster (1989) and, associated with this perspective, the role of temporal causation (Stinchcombe 1968; Pierson 2004; Falleti 2005; Mahoney and Goertz 2006; Mahoney 2008; Grzymala-Busse 2010; Mahoney and Thelen 2010; Bennett and Checkel 2015; Falleti and Mahoney 2015). The most significant advances in tools for data collection can be found in the development of protocols for the systematization and codification of semi-structured in-depth interviews (e.g. Campbell et al. 2013). Finally, in recent years, some scholars have started to talk about a transparency revolution in qualitative research (Moravcsik 2012, 2014a, 2014b; Büthe et al. 2015). Although there have been significant methodological contributions previously, the progress in the last decade or so has been impressive.

Nonetheless, there are still several areas where more improvement is needed, especially in the link between theoretical development and data (and the return to theory)—that is, the iteration between theory and data. This research note aims to help advance analytical transparency: to clarify the connection between conclusions, evidence and the analytical assumptions and decisions that lead to a specific interpretation of the data (Lupia and Elman 2014; Büthe et al. 2015).

In the context of the experimental revolution occurring in several social science disciplines, social research scholars and organizations are promoting the design and registration of pre-analysis plans (hereafter abbreviated as PAP) to ensure transparency. In a PAP, the researcher registers the analysis and tests that he will conduct with the data. This procedure aims to reduce the temptation to fish for statistically significant relationships that have not been explicitly considered in the research design. In political science, for example, one of the most interesting initiatives is one promoted by the network Evidence in Governance and Politics (EGAP). In this paper, we seek to extend the practice of the design

---

3 However, the concern over transparency predates this paradigmatic revolution. King (1995), for example, already insisted on this issue.
and registry of PAP to qualitative research: PAP-Q. For those unfamiliar with PAP registration, it is important to highlight that neither the PAP nor the PAP-Q, here developed, differ from what a professor of social scientific methodology would expect from a student’s research design. The main differences between a classic research design and a PAP (or PAP-Q) are, essentially, twofold: first, the PAP is explicitly public and easily accessible to other researchers; and second, the PAP commits the researcher to a particular analysis of his or her data. Both features of a PAP seek to improve transparency in social research. At this point, it is important to highlight that within the qualitative tradition many forms of research use some form of registration. In the constructivist tradition, the use of a notebook is a way to register the learning process in real time (Geertz 1973). This, to some degree, inhibits confirmation bias because it is registered live—the researcher cannot rationalize a finding quickly enough for the field notes to reflect that ex-post rationalization.

II. THE REASONS TO USE A PAP-Q

Why is a PAP necessary in qualitative research? How can this kind of tool be justified in a type of research that is essentially inductive? Both questions are intimately related, and the answer to both questions is generally the same: research, by definition, must be a systematic process. In a column published in *The Political Methodologist*, Andrew Moravcsik wrote:

> The unique interpretive flexibility qualitative scholars enjoy only serves to increase transparency requirements. Any appropriate and workable standard of qualitative transparency must be suited to this type of research. It must preserve the basic narrative “process-tracing” structure of presentation. Scholars should provide readers with the data and analytical interpretation of each piece of evidence in context, and a methodological justification for their selection. Readers must be able to move efficiently, in real time, from a point in the main narrative directly to the source and its analysis, and back again—a function traditionally carried out by footnotes and endnotes. Third, analytic transparency provisions must permit scholars to explain the interpretive choices they have made with regard to each piece of evidence. All this must take place within the real-world constraints set by intellectual property law, human subject protection, logistics, first use rights and existing publication formats (Moravcsik 2015).

In fact, all the advances briefly reviewed earlier in the introduction aim at improving research systematization. The first edition of *Rethinking Social Inquiry* (Brady and Collier 2004) agreed with KKV that this was a key aspect for any research enterprise, and the authors acknowledged qualitative research had a long way to go to improve in this aspect. The consensus on shared standards...
for systematic research thus accelerated the process. For example, the process-tracing tests developed in the last decade provide clear-cut parameters for hypothesis testing in comparative historical analysis, thus leaving behind the notion that any well-told story constitutes good process tracing. The proposal for a PAP-Q aims to continue improving in this regard.

The PAP-Q takes on the idea that a great proportion of qualitative research work (and much of its virtue) lies in its inductive character. Nonetheless, its inductive nature does not preclude the development of theoretical claims, and does not entail that everything be learned or done in the field or through the compilation of secondary sources. The PAP-Q seeks to establish a formal beginning of the iteration between empirical work and theory—a frontier for the iteration between theory, evidence and the interpretation of the evidence that is typical of, for example, a comparative-historical study (Mahoney and Rueschemeyer 2003b).

The PAP-Q also allows for the fine-tuning of the connections between conceptualization, measurement and analysis that still operate as separate entities in social science research. In fact, this lack of dialogue between such components also applies to quantitative social science, whether observational or experimental. Even though some scholars have highlighted the need to carefully work with the concepts applied in measurement (Goertz 2006; see also, Sartori 1970) and others have provided improvements in the use of analytic tools or have developed protocols for systematizing qualitative information, there are no works to date that connect these efforts and suggestions. The proposal for a PAP-Q seeks to take one step further in the articulation of these initiatives to improve qualitative research in all its stages. For that purpose, we take advantage of advances in experimental research.

The PAP-Q orders the protocol before beginning the immersive processes of fieldwork or archival research. It also orders the iteration between theory-data-theory that characterizes qualitative research. What are a researcher’s expectations in terms of the causal process observations (Collier et al. 2010b) he or she expects or needs to encounter? Perhaps the earliest study to implement this logic was (Fairfield 2013). In the appendix, Fairfield lists the hypotheses, observations and inferences using Van Evera’s (1997) process tracing tests. The PAP-Q seeks to concentrate more effort in the design stage of a project in order to clearly establish the characteristics of the expected evidence in light of the initial theoretical formulation. In the relationship between theory and evidence, the PAP-Q forces the researcher to think about the design of the project and about ways to avoid confirmation bias. In the data collection stage, it stimulates

---

4 It is very interesting that this process experienced by the positivist tradition of qualitative research, at least in political science, was accompanied by a fluid dialogue with experimental social science. An example of this dialogue can be observed in the CQMR summer school and in its Southern Cone counterpart held at the Instituto de Ciencia Política of the PUC-Chile.
the researcher to dedicate more time to list the greatest possible number of empirical manifestations that he thinks he will need to find in the field.

We hope that researchers read this proposal as a way to avoid the conventional tension between in-depth knowledge, analysis and theory building. Moreover, the PAP-Q allows other colleagues to understand what initial assumptions were confirmed or disconfirmed in the empirical work, as well as what theoretical claims emerged from fieldwork and engendered changes in the theory. In this vein, the PAP-Q also functions as a research log—a tool that is deeply ingrained in the non-positivist qualitative tradition.

III. THE CONTENTS OF A PAP-Q

Experimental research in the social sciences does not have a standardized format for PAP. As already mentioned, EGAP has established a format with parameters and very general guidelines for the registry of trials. Nonetheless, there is no disciplinary convention as exists, for example, with respect to the value of statistical significance. At the same time, the specific features of qualitative research (especially its inductive nature) force us to significantly adapt existing models of PAP that predominate in experimental research.

The sections and contents of our PAP-Q model include the recent (i.e., in the last ten years) advances in the systematization of qualitative research. What now follows is no more than a first effort to establish basic parameters of a PAP-Q. Rather than “set in stone” its components, our main goal is to stimulate the development of this practice as a positive habit in qualitative research.

Conceptualization

In 1970 and in the context of an expanding political science, Sartori highlighted the need to take good care in the use of concepts, noting that conceptualization necessarily precedes measurement. Years later, Goertz (2006) strengthened this principle in his manual on concept formation. At the same time, the various methodologists who have elaborated the group of manuals that form the revolution in positivist qualitative research (e.g. George and Bennett 2005) have insisted that one of the greatest virtues of this type of research lies in its greater potential for concept and theory building. Thus, it seems reasonable that a

---

5 For further details, see http://egap.org/design-registration-instructions.
6 The PAP-Q should be registered before the researcher begins fieldwork or the collection of primary or secondary sources for historical analysis. In the absence of a repository, it will suffice for now that every scholar post the PAP-Q on his or her professional webpage.
first section of the PAP-Q be devoted to the discussion and justification of the concepts used in one’s research.\textsuperscript{7}

The questions that the researcher should answer in this section are highlighted by Goertz (2006): What is the dependent variable? What are its attributes? What kind of empirical referents (indicators) operate as manifestations of it? Also, how do attributes and indicators of the concept relate? In this vein, Adcock and Collier’s (2001) article on measurement validity provides a key point for this section of the PAP-Q. In the PAP-Q, scholars can work on differentiating between the “systematized concept” and the “background concept” (Adcock and Collier 2001: 531); i.e., scholars can work on the formulating their concepts by relying on a “constellation of [available] meanings” to arrive at a systematized concept. Through fieldwork, they can then refine and revisit the background concept. The PAP-Q, then, can help first with conceptualization and with describing the process of measurement, while fieldwork can help boost the initial conceptualization (with modifications that should be registered in amendments to the PAP-Q).

**Theory: Inductive and Deductive Section**

The theoretical section is crucial in any PAP, and a PAP-Q is no exception. In qualitative research, the researcher develops his theory in interaction with the empirical work. Nonetheless, qualitative researchers of the positivist tradition begin their studies with a theoretical expectation and with their arguments already developed to a certain degree. We thus propose that this section of the PAP-Q consist of two sub-sections: deductive and inductive.

One of the main trade-offs in qualitative research is that between parsimony and complexity (Collier, et al. 2010a). Immersion in the field and the accumulation of evidence and details that pay tribute to the in-depth nature of qualitative work are usually in tension with the elaboration of parsimonious arguments. Moreover, and as in any other observational study, the researcher has to deal with his own confirmation bias (see Dunning 2015). We will return to this issue in the design section. The deductive section consists of the researcher’s discussion with specialized literature for the research problem that he is dealing with, and it is the section in which he formulates his initial theoretical claims. This section thus seeks to delineate the theoretical claims as clearly as possible before the complexity of the field invades one’s neat formulations.

The inductive section (the registry of iterations between the theory and the empirical work) is meant to detail any changes in theoretical understanding and the subsequent changes made to the deductive section. The registry of this

\textsuperscript{7} This is also valid for experimental research, as experimental PAPs should also include a section on conceptualization.
section could be conducted after finishing the accumulation of evidence (e.g. fieldwork with interviews or the revision of primary or secondary historical sources). However, it is appropriate and preferable to register the moments of change (iterations) throughout the research process itself, in order to avoid later theoretical rationalizations. Thus, this section of the PAP-Q has two purposes: first, as with every research design, it seeks to order the job for the researcher; and second, it distinguishes the initial theoretical argument from what has been learned through later fieldwork. In this way, the final result is better understood, as is the reasoning of the researcher, thereby increasing the possibility of replicating the research.

In the inductive section, therefore, the researcher should clarify how, why and even when he identified a given factor as a constitutive part of the theoretical argument. Through the registry of the iterations between theory-data-theory, the goal is to publicize how the researcher theoretically rationalized the accumulated evidence and what he learned from the fieldwork.

**Design**

A. *Working Hypotheses*: One of the main virtues and distinctive features of qualitative research is its emphasis on non-linear causation (Goertz and Mahoney 2012). In this tradition, scholars refer to multi-causality, conjunctural causation and mechanistic causation (Mahoney and Rueschemeyer 2003a; Bennett and Checkel 2015; Mahoney and Thelen 2015). In this vein, in the working hypotheses sub-section of the PAP-Q, the researcher should develop these notions and answer the following questions:

- **Complex causation**: What are the interactions between explanatory factors? Is there a relationship of necessity and/or sufficiency between the theorized causes?
- **Mechanistic causation**: What role does the researcher assign to the order of causal factors? What is the role of time in this explanation? What kinds of causal mechanisms are expected to result in observable empirical manifestations?

As in the *Theory* section, the researcher should update his hypotheses indicating how the evidence engendered changes in his working hypotheses, whether by the inclusion of new factors or by changing the relationships among them. In the first registration, the researcher should register both his main hypotheses (in as much detail as possible) as well as the alternative hypotheses. Given the inductive nature of qualitative research, the main hypotheses and the alternative hypotheses will be updated in future iterations. At this stage in the research process, the alternative hypotheses should be conceived as plausible counterfactuals.
B. Case Selection: One of the most significant advances in the systematization of qualitative methods has been the formalization of the criteria used for case selection and the identification of different types of case-centered studies (Eckstein 1975; George and Bennett 2005; Gerring 2006). Thus, in the PAP-Q, the researcher should determine the type of case study research that he will conduct. Also, Gerring (2006: 19-20) discusses the need to delimit case studies in temporal and spatial terms, and he emphasizes that the biggest challenge resides in the temporal delimitation of one’s case study: “We know, more or less, where a country begins and ends, while we may have difficulty explaining when a country begins and ends.” Hence, it is important to state from the outset the justification of case selection, the type of case study to be conducted and the delimitation in terms of space and time.

C. Tools of Data Collection: There are multiple tools for data collection, and for the purpose of this proposal we will detail the issues that a researcher should include in a PAP-Q whenever he works with semi-structured, in-depth interviews or when his fieldwork consists of archival historical sources. In the first case, in the PAP-Q, the researcher should define the questionnaire beforehand in as much detail as possible and relate it to the registered hypotheses. Also, the PAP-Q should establish the criteria for the theoretical sample of interviewees and, ideally, provide a list of potential interviewees. Listing the prospective interviewees is an exercise through which the researcher gives an idea of the type of actors from whom he seeks to obtain information or discursive content to test his theory—that is, to observe empirical manifestations of a theorized causal mechanism. The researcher should also indicate the repository where he will deposit the transcripts. However, in doing so the researcher should take into consideration the degree of confidentiality and anonymity agreed upon by each interviewee as per the principle of informed consent. We see this task as an advantage of the PAP-Q and not simply a requirement to comply with a standard of research transparency. Moreover, pre-registration of the deductive portion of the theoretical argument encourages one to reflect upon outcomes and thus motivates us to design better measurement tools before we spend resources in the field. Pre-registration of theory and questions also encourages reflection concerning what type of individuals to interview and how each interviewee relates to the theoretical argument.

One of the most repeated recommendations in the qualitative methods manuals is to heed the inherently in-depth nature of qualitative research. In the case of historical sources, George and Bennett (2005) highlight the need to revise more than one historiography. Indeed, in the PAP-Q the researcher should clearly state the historiographies that he will revise in his study—which certainly does not imply that he must limit himself to those sources. Once the fieldwork or the data collection stage is over, the researcher should update this information

8 For archiving in-depth interviews, see the work of Elman et al. (2010). A good available alternative is the Qualitative Data Repository of Syracuse University.
in the PAP-Q, explicitly indicating the additions or changes with respect to the original registry and the reasons for such changes.

D. Analysis: A virtue of PAP registry in experimental research is that it seeks to avoid the bias of presenting statistically significant results at any cost, and to guide the incentives towards the design. This spirit is perfectly replicable in qualitative research. We insist that scientific inquiry, whatever its specific features in a given study, is above all a systematic process. If the researcher decides to conduct an analysis of in-depth semi-structured interviews through content analysis, he should previously register, in the most detailed possible way, the criteria he will follow to codify them. In qualitative research there have been many advances in terms of standards for codification. A very useful guide can be found in the work by Campbell et al. (2013). We believe that doing this exercise before fieldwork can stimulate the researcher to develop a well-conceived interview instrument, reducing the risk that the researcher might later regret not having developed good interviews that he might no longer be able to conduct. Thus, the goal of this section of the PAP-Q is to stimulate the habit of developing well-conceived measurement instruments.

In the case of comparative historical analysis, it is appropriate that the researcher derives from his hypotheses the type of tests that he expects to conduct in the fieldwork. Nonetheless, there is also room for the understandings and rationalizations achieved throughout the empirical work, which should be registered. Thus, on the inductive side of comparative historical analysis, the researcher has to make explicit claims about the discovery process and the rationale for the inclusion of his causal process observations (Collier et al. 2010b).

Before closing, it is important to state that several methodologists from different strands and epistemological perspectives of the qualitative tradition have been developing sets of concrete good practices that operate as protocols to promote transparency in the analysis and replication of the evidence (see Büthe, et al. 2015). As we highlighted in the beginning of this paper, we have placed more emphasis on detailing the grey areas of what could constitute a PAP-Q. A great example of a detailed form for registration is provided by Gleditsch and Kern (2016).

IV. CONCLUSIONS

In this research note, we have proposed the development of PAP-Q in the social sciences. With this mechanism, we first tried to systematize and unify the various advances of the qualitative positivist research tradition in terms of design, theory building, case selection, transparency and the systematization of qualitative data. The connection between theory development and evidence is still fragile in this research tradition. Thus, beyond synthesizing these advances in one document, the PAP-Q strengthens the connection between these components of a research enterprise. For that purpose, we used the
experimental social science model of PAP and we adapted it to the inherently inductive character of qualitative methods.

Moreover, beyond reinforcing the connection between theory and evidence, and encouraging greater transparency in qualitative research, the PAP-Q supports and guides the work of the researcher by structuring the process and by pushing the researcher to carefully think of the design. The PAP-Q does not seek to impose a corset that chokes creativity and prevents the emergence of casual findings; rather, the PAP-Q should stimulate creativity through design and enable the researcher to be open to finding evidence that runs against the researcher’s prior expectations or that enables us to rethink our theories. In qualitative studies, researchers might be suffocated by the abundance of elements to consider and by the predisposition, due to confirmation bias, to consider only certain types of evidence (or to read it in a predetermined way). The PAP-Q seeks to inoculate the researcher against these types of problems.

Finally, the importance of registering the process of iteration between theory and empirical data could well be extended to PAP in experimental research. Some research projects (e.g. Dunning et al. 2015) register amendments to their original PAP. In such amendments, the authors update modifications to their theory or to the empirical strategy in light of the learning process that takes place during the research though, of course, prior to the analysis of the data. Thus, while this practice already exists, it is not commonly used. Nonetheless, the implementation of field experiments is full of complications, and the researcher should modify or calibrate the intervention as needed. This also includes updating his theoretical expectations. One might think that one is manipulating a given variable, but one can modify and make more precise one’s expectations based on knowledge of the field where one will conduct the intervention. Thus, researchers should register, in the most reasonable and feasible way, the updates and learning that result from implementing an intervention, so that future implementations of a similar intervention (or in the same setting) take those lessons into consideration and can benefit from them.

REFERENCES


9 A quick review of registered PAPs on the EGAP web page shows how limited this practice is at present.


Moravcsik, Andrew. 2012. “Active Citation and Qualitative Political Science”. *Qualitative and Multi-Method Research* 10(1): 33-37.


---

**Rafael Piñeiro**: Assistant Professor, Departamento de Ciencias Sociales y Políticas, Universidad Católica del Uruguay. Correo electrónico: rafael.pineiro@ucu.edu.uy

**Fernando Rosenblatt**: Assistant Professor, Escuela de Ciencia Política, Universidad Diego Portales, Chile. Correo electrónico: fernando.rosenblatt@udp.cl