

Iteration in Mixed-Methods Research Designs Combining Experiments and Fieldwork¹²

Verónica Pérez Bentancur
Department of Political Science
Universidad de la República, Uruguay

Lucía Tiscornia
Division of International Studies
CIDE, Mexico

Forthcoming, Sociological Methods and Research

Abstract

Experimental designs in the social sciences have received increasing attention due to their power to produce causal inferences. Nevertheless, experimental research faces limitations, including limited external validity and unrealistic treatments. We propose combining qualitative fieldwork and experimental design iteratively—moving back-and-forth between elements of a research design—to overcome these limitations. To properly evaluate the strength of experiments researchers need information about the context, data, and previous knowledge used to design the treatment. To support our argument, we analyze 338 pre-analysis plans submitted to the Evidence in Governance and Politics repository in 2019 and the design of a study on public opinion support for punitive policing practices in Montevideo, Uruguay. The paper provides insights about using qualitative fieldwork to enhance the external validity, transparency and replicability of experimental research, and a practical guide for researchers who want to incorporate iteration to their research designs.

Keywords: iteration, research design, experiments, qualitative fieldwork, mixed-methods

¹ Authors in alphabetical order, equal contribution assumed.

² Part of the reflections in this article resulted from our experiences developing a research project with Juan Albarracín and Leslie MacColman, funded through the Agencia Nacional de Investigación e Innovación (ANII, Uruguay), project number FSSC_1_2018_1_147720.

1. Introduction

Social science researchers frequently discuss how to improve causal inference— why and how something occurs— to better understand the inner workings of social life. This concern has motivated efforts to improve research methodology. As part of these efforts, experimental designs have become increasingly popular in the social sciences (Druckman and Green 2021). In general, experiments are research strategies that can be used for descriptive and causal exploration (Blair, Coppock and Humphreys, forthcoming). For example, list experiments can be used as tools to document stages of a process to fill in evidentiary gaps (Gonzalez-Ocantos and LaPorte 2019), and randomized controlled trials (RCTs) can be used to infer causality (Dunning 2012; Gerber and Green 2012; Blair, Coppock and Humphreys, forthcoming). Nevertheless, critics argue that randomization has limits, and that experiments typically lack generalizability and are often unrealistic. A nascent body of literature highlights the advantages of using qualitative tools to refine experiments in the design phase as a tool to overcome these limitations (Auerbach and Thachil 2020; Beach and Littvay 2020; Dunning 2008, 2012; Dunning and Harrison 2010; Thachil 2018; Seawright 2016, 2021).

We build on this body of work and make three contributions: 1) we clarify what is entailed in building realistic treatments and specify how qualitative tools can contribute to this goal; 2) we describe how qualitative methods can be used to improve measurement during the experimental design phase; 3) we propose iteration as a specific procedure to improve measurement by operationalizing untested assumptions in experiments (Seawright 2016, 2021). Iteration—the process of alternating between different elements of the research design (Fairfield and Charman 2017; Kapiszewski, MacLean, and Read 2015)— is a means to combine qualitative fieldwork and experiments. Using our own research as an example, we provide concrete steps on how to mix these two methods applying iteration in the design phase.

Iteration is an important tool for at least three reasons: First, it improves causal identification. Iteration between theory and fieldwork produces a more refined understanding of the context and more realistic interventions—those that more accurately depict real-life situations (Seawright 2021). Second, when made explicit, iteration allows the research design to be more transparent. By outlining the iterative steps, researchers facilitate replication. Third, transparency about iteration aids generalization: if readers are told how a research design was informed by or tailored to features of the study context, then they are in a better position to understand which aspects of the findings are more or less likely to travel to other contexts. Gaining a deeper understanding of the context can clarify the scope of the theory being tested, and thus provide insights regarding the extent to which findings can be generalized.

Our approach is different from using qualitative fieldwork in the analytical phase of the research process, typically employed to discover causal mechanisms (Dunning 2015; Levy Paluck 2010). Emphasizing iteration can help researchers structure their projects before carrying out the analysis. Iteration is also different from pre-testing the experimental design. While pre-testing focuses on piloting the mechanics of a design on a smaller sample prior to fielding a study, it may or may not allow researchers to discover whether the elements of the design reflect the context.

The paper is organized as follows: We first review scholarly work combining experiments and qualitative evidence and we describe the process of iteration with qualitative fieldwork in experimental research designs. We show that the authors of experimental studies that mix methods, rarely make the steps they follow explicit. To illustrate this point, we analyze 338 pre-analysis plans submitted to the Evidence in Governance and Politics (EGAP) repository in 2019. Then, we illustrate our proposed iterative process by describing a study of public opinion support for punitive policing practices in Montevideo, Uruguay, that

combined a survey experiment and qualitative fieldwork. Finally, we summarize our contributions regarding the value of iterating between qualitative fieldwork and experiments, we discuss some limitations of this approach, and we include a practical guide for researchers who wish to incorporate iteration in their research designs.

2. How to iteratively mix methods in the context of an experimental design

Social science research seeks to provide accurate causal inferences to advance understanding of social, economic and political phenomena. In recent years, this concern has manifested in a growing interest in improving several aspects of the research process, including, among other issues, improving research designs (Brady and Collier 2010; Gerring 2011; Goertz 2017; King, Keohane, and Verba 1994; Lieberman 2005; Seawright 2016) increasing transparency in qualitative research sources (Lupia and Elman 2014) and in quantitative and qualitative designs through pre-registration processes (Blair et al. 2019; Elman, Gerring, and Mahoney 2020).

Social science researchers in the quantitative tradition have increasingly used experimental designs, including field experiments, lab experiments, natural experiments, and others. Though initially marginal, experimental designs are widely used, owing to their particular strength at producing causal inferences (Gerber and Green 2012; Jackson and Cox 2013; Hainmueller, Hopkins, and Yamamoto 2014). Experiments derive their potency from two features: the combination of a relatively large number of cases with random assignment to treatment and control groups. Random assignment to treatment is meant to eliminate systematic group differences at the outset of the study such that any observed difference between treatment and control groups in the outcome variable can be attributed to the treatment of interest. The main advantage of experiments for producing strong causal inferences thus lies in their ability to eliminate the effects of confounding variables present in

most observational designs in both quantitative and qualitative research (Dunning 2012; Gerber and Green 2012). Experiments can also support causal inference by demonstrating the plausibility of a certain causal path by filling in evidentiary gaps in cases where mechanisms are hard to observe directly (Gonzalez Ocantos and LaPorte 2019: 16-18). Nevertheless, scholars have pointed out that whether experiments are the best tool for causal inference depends on the questions we are asking, as well as prior knowledge (Deaton and Cartwright 2018: 2).

In this paper we focus on two limits of experimental research: unrealistic treatments and lack of external validity (Blair and McClendon 2021; Dunning 2015; Seawright 2016). Realistic interventions are such that participants could be expected to reproduce real-life behaviors. Constructing real-life scenarios requires considering the context to identify interventions that are adequate reflections of the context (Seawright 2016:166-167; 2021:371). Realistic interventions are central to external validity. Replicating an experimental design requires adapting it to a new context (Deaton and Cartwright 2018; Seawright 2021). We argue that to overcome these limitations researchers can use a mixed-methods approach that explicitly incorporates qualitative elements through an iterative process in the design phase. As a result, researchers improve the theory and refine the experimental intervention, which aids the inferential power of the argument. An added benefit of making the iterative process explicit is increased transparency of the design process, which aids replicability as well as generalizability.

A mixed-methods approach involves combining strategies for data collection and analysis from different methodological schools, usually from quantitative and qualitative approaches (Goertz 2017; Lieberman 2005; Seawright 2016). Combinations of methods from different epistemological traditions are multiple: for example, surveys can be combined with interviews, focus groups, or process tracing; regression analyses of a large number of cases

can be combined with an in-depth study of one or two cases; formal theory can be combined with case studies, among other possibilities (Cyr 2019; Dunning and Harrison 2010; Lieberman 2005; Levy Paluck 2010; Redlich Revkin 2020; Seawright 2016). Mixed methods can be used to probe the untested assumptions of other methods, such as assumptions about measurement validity or about the nature of data (Seawright 2016: 48-55).

A central distinction is between mixed-methods and the process through which methods are mixed. “Mixing methods” refers to the combination of quantitative and qualitative methods.³ Different methods can be combined through different processes: nested analysis (Lieberman 2005), parallel (Harbers and Ingram 2020), embedded (Hollstein 2014), integrative (Seawright 2016), iterative (Kapiszewski, MacLean, and Read 2015). While nested, parallel, embedded and integrative techniques are generally employed in the *analytical phase* of the research process, we argue that iteration is particularly useful in the *design phase*. Iteration is one possible way to combine methods. It refers to the alternation between different elements of the design (for example, using in-depth interviews to refine a survey measurement instrument and then validating it with additional interviews), thus making the researcher more conscious of each method’s contribution to inference.

Related, another relevant distinction is between integration and iteration. Integration refers to the combination of two or more methods to support a unified causal inference, which corresponds to the analytical phase of the research design (Seawright 2016). Iteration, we contend, does not necessarily involve an analytical component, it focuses on strengthening the design prior to implementation and analysis (Kapiszewski, MacLean, and Read 2015).

³ Qualitative methodology encompasses a wide variety of tools with varying degrees of involvement and difficulty (in depth ethnography, interviews, review of press, focus groups, among others). Ultimately, the method selected will depend not only on the problem to be addressed, but also on considerations about time and resources.

Iteration is generally understood as the movement back-and-forth in the research process. There are different ways to conceptualize iteration: for example, between methods of reasoning (induction and deduction), or between different elements of the design itself (theory and fieldwork) (Kapiszewski, MacLean, and Read 2015; Yom 2015). Iteration has a dual objective: the refinement of theory, and the refinement of the research design itself (Dunning 2008; Fairfield and Charman 2017; Kapiszewski, MacLean, and Read 2015; Yom 2015). Through iteration researchers can revise and update different parts of the design: the research question, concepts, hypotheses, cases and instruments for data collection, such as interview protocols and survey questionnaires (Kapiszewski, MacLean, and Read 2015: 23). Iteration is also a mechanism to improve measurement because it allows to probe the untested assumptions of different methods, rectifying when necessary (Seawright 2016). For example, an experiment designed only following an extensive review of the literature likely makes several assumptions about measurement and face validity.⁴

In some cases, the distinction between iteration in the design and analytical phases may not be so clear cut. For example, in designs that implement process-tracing or comparative historical analysis, researchers may iterate with insights from the analytical phase to refine their hypotheses and theory (Yom, 2015). However, when designs include components that are “one shot” such as the implementation of a survey, or an experiment, the distinction between iteration in the analytical and design phases is more clearly defined (since the experiment cannot be redone once it is fielded)⁵. This is also possible in qualitative designs, for example, those using process-tracing, where iterating with interviews can lead to

⁴ We are grateful to an anonymous reviewer for this insight.

⁵ In process-tracing designs, some interviews are also “one shot”. However, because interviews are sequenced, the researcher does not conduct all interviews at once and can make changes to the interview protocol. In addition, different sources of information can inform the same piece of evidence (González-Ocantos & Masullo 2020).

theory refinement and modifications in interview questionnaires that allow for gathering better pieces of evidence (González-Ocantos & Masullo 2020).

Despite being a common research practice, iteration is often unacknowledged, as it is viewed as pre-scientific (Beach and Littvay 2020; Yom 2015). We argue that when iteration is made explicit and public it makes the research process more transparent, as it requires one to make all the steps in the design known. This paper's contribution lies in specifying how to iterate between methods—in particular, the combination of qualitative fieldwork and experimentation—to improve a research design prior to testing with the goal of strengthening causal inference. In doing so, we bring to bear mixed methods research containing experiments, research that mixes methods in the design phase without making it explicit, and methodological research that explicitly discusses iteration.

Although there is a growing body of research that mixes experiments with qualitative fieldwork (for some examples, see Auerbach and Thachil 2020; Clayton et al 2020; Dunning 2008, 2015; Dunning and Harrison 2010; Levy-Paluck 2010; Rao, Ananthpur and Malik 2017; Thachil 2018), the process of mixing methods is not always acknowledged. Within those researchers who mix methods, a common practice is to incorporate qualitative methods during the analytical phase, to illustrate causal mechanisms. For example, Levy-Paluck (2010) deployed a survey that had an experimental component as well as open ended questions. Using the information from the open-ended questions to contextualize the experimental results, the author concluded that combining experiments and fieldwork produces a more complete picture of the causal effects resulting from the experiment (Levy-Paluck 2010: 61). Other scholars might use qualitative fieldwork to strengthen the design but without making it explicit (Beach and Littvay 2020), and methodological research highlighting the importance of iteration in the research process places less emphasis on its use to improve experimental designs prior to testing (Kapiziewski et al. 2015; Yom 2015).

We build on previous approaches highlighting the value of putting fieldwork at the service of experiments to improve the latter's ability to yield valid and powerful causal inferences (Dunning 2008; Kapiszewski, MacLean, and Read 2015; Seawright 2016). Specifically, we identify and stylize an iterative sequence for refining a research design that can be applied more generally. In addition, we contend that making iteration explicit forces researchers to specify every step of the design, thus contributing to transparency and increasing potential generalizability.

Using iteration in mixed-methods designs helps researchers make stronger causal inferences than they would have been able to make otherwise. Combining methods through iterative steps allows researchers to leverage the strengths of both quantitative and qualitative approaches in the design phase. However, iteration is not an indefinite process. In a design containing an experiment, iteration ends when the experiment is fielded. To make decisions about the exact duration of the iterative phase, scholars should consider substantive and pragmatic criteria. Substantively, iteration ends when qualitative fieldwork does not reveal new insights that would significantly improve the experimental design. This is akin to the notion of "saturation" in qualitative research: when conducting interviews or participant observation, the researcher does not gain additional confirmatory evidence, or new evidence that challenges their argument (Saunders et al. 2017). Pragmatically, when to end iteration is conditioned by the amount of resources available to the researcher (time, and economic resources) (Seawright and Gerring 2008). Other pragmatic considerations relate to safety and security conditions on the ground (Chappuis and Krause 2019).

The benefits of iteration

In this section we describe a few examples of research designs that illustrate some of the benefits of iteration. It is worth noting that not all mixed-methods research using

experiments requires iteration. For example, a common way of mixing methods to generate designs are natural experiments. In natural experiments, qualitative methods are a core element used to produce the design; their key function is to provide evidence of the pseudo-random nature of treatment assignment. Archival research, review of documents, interviews, and other qualitative tools provide in-depth knowledge of cases that show random assignment of treatments in the real world. Qualitative methods are necessary to show the mechanisms whereby treatment and outcome are connected (Dunning 2012; Ferwerda and Miller 2014; Kocher and Monteiro 2016; Seawright 2016).

One example that illustrates the importance of iteration in experimental contexts is Clayton et al. (2020). The authors conducted a survey experiment to assess whether women candidates suffer gender bias in Malawi. In parallel with the survey, they gathered biographical information about the candidates and conducted focus groups with them. Initially, they aimed to use the qualitative data to contextualize the experimental results. However, their main hypothesis—that in conservative contexts citizens prefer male candidates—was not supported by the experimental results, making it difficult to align the survey findings with the qualitative evidence. In reviewing the latter, they realized that the candidate profiles they used in the experiment were not entirely realistic. A relevant finding from the focus groups is that women candidates in Malawi face defamation campaigns, which impacts citizens' relative evaluations of female and male candidate profiles. The authors conclude that this contextual information was not captured in the experiment, and that this can explain their null finding. They also conclude that they would have benefitted from conducting the focus groups before the experiment, to improve the design of the latter by including a “‘rumor mongering condition’ to test whether and how voter biases are activated in the electoral process” (Clayton et al. 2020: 622). This is an example of fieldwork conducted in parallel with the implementation of the experiment and used mainly for

analytical purposes. Had the authors implemented iteration between fieldwork and the experiment in the design phase, their experiment would have been stronger and their findings might have differed. Similarly, Rao, Ananthpur and Malik (2017) conduct extensive ethnography in parallel with a survey-based RCT. They use the ethnographic fieldwork analytically to unpack the reasons for the failure of the RCT and to illustrate small treatment effects that would not have been detected through statistical analyses.

An example that uses ethnographic methods for purposes of strengthening the design of a conjoint experiment is Auerbach and Thachil (2020), who conducted research on slum leaders in India. To create realistic vignettes, they use ethnographic fieldwork to define accurate experimental attributes (Auerbach and Thachil 2020: 476-477). The use of ethnography results in a compelling research design with realistic treatments. However, the authors do not specify the concrete steps they followed to improve the treatment. From the paper it is unclear whether they designed the experiment first, and then iterated with fieldwork to adjust the original design, or they sequentially conducted fieldwork, and designed the experiment.

Explicit examples of the benefits of iteration are Thachil (2017, 2018), Dunning and Harrison (2010), and Jha, Rao and Woolcock (2007), who conducted experiments to understand identity politics in India, ethnic cleavages in Mali, and the construction of local leadership in India, respectively. Thachil (2017, 2018) uses ethnographic fieldwork with the dual purpose of defining his sample and to improve a vignette experiment. The original experimental intervention was adjusted using information from fieldwork to better reflect the context. In turn, Dunning and Harrison (2010) constructed a preliminary matrix of common Malian last names—as proxies for ethnicity—based on interviews, to assign subjects to the different treatment conditions. They then validated and refined the original matrix with a second round of interviews with qualified informants in the field. This process allowed them

to add more last names to the matrix. Prior to conducting the experiment, they tested the revised matrix with 169 people, which allowed them to further refine it (Dunning 2008: 21-22). Using a participatory econometrics approach, Jha, Rao and Woolcock (2007: 231-232) constructed survey instruments based on focus groups discussions, in-depth interviews, and participant observation, which they then pre-tested and adjusted through additional interviews.

The examples above illustrate how research designs containing an experiment can benefit from the use of iteration between qualitative data collection and the design of the experiment prior to implementation. The examples also highlight how the absence of iteration can lead to poor causal identification strategies. In the following section we present a more systematic analysis by evaluating a set of 338 pre-analysis plans.

3. How common is iteration in experimental designs?

The process of combining methods and iterating between different parts of the design in experimental contexts is rarely discussed. Thus, researchers have little guidance on how to iterate or about why making the process explicit is valuable. Although experimental designs rely on fieldwork, interviews, and case knowledge, this fact is rarely made explicit or public. Yet the strength of an experiment can only be properly evaluated if researchers know what kind of data and prior knowledge went into its design.

To illustrate this point, we analyze 338 pre-analysis plans (PAPs) pre-registered in the Evidence in Governance and Politics (EGAP) repository in 2019 (out of a total of 457 pre-registered designs) for which documentation was available.⁶ EGAP is a cross-disciplinary network of researchers and practitioners focused on experimental designs.⁷ EGAP members'

⁶ For 119 of the registered designs, documentation is either gated or there is no additional documentation available.

⁷ See here: <http://egap.org>

research designs encompass a variety of topics, such as inequality and poverty, governance and institutions, conflict and violence, among others, across disciplines such as sociology, political science, economics, public health. These designs are not limited to program evaluations, they also include academic research. PAPs can be pre-registered before conducting the research, or prior to looking at the data, after pre-testing. EGAP's repository is one of the most flexible in terms of what can be pre-registered (Boudreau 2021). Unless they are gated, these documents are available to the public.

In their simplest form PAPs contain the basic structure of the research design: hypotheses, the structure of the experiment, the population to which it will be applied and the analytical strategy. The main goal of pre-registration is to prevent analytical biases, such as ad hoc analyses ex post, or choosing which results to report (Blair et al. 2019; Jacobs 2020). Another goal of pre-registration is related to increasing research transparency⁸. Transparency in every step of the research design makes replication possible (Lupia and Elman 2014; Pérez Bentancur, Piñeiro Rodríguez, and Rosenblatt 2018). An added value of pre-registration, much less discussed or acknowledged, relates to making the process of design itself public. Many studies are not published due to a wide variety of biases in the publication process, and a strong reluctance to publish null results (Jacobs 2020). In this context, pre-registration is key because it allows other researchers to learn from the knowledge that is generated in the process of designing research.

We chose PAPs for pragmatic and substantive reasons: pragmatically, PAPs are centralized in one repository, with the same guidelines, containing a diversity of projects across disciplines. More importantly, there is a powerful substantive reason to focus on PAPs: they contain explicit decisions related to the design phase of the research process, which is our focus. Therefore, if the different steps in the construction of the design are

⁸ This is explicitly stated in EGAP's statement of principles.

specified, and if there is iteration between those steps, we should expect to find evidence in PAPs. An alternative analytical strategy could have been to select articles published in academic journals. However, conducting this selection process with journals poses challenges in terms of the substantive requirements, variation across disciplines and journal access. In addition, had we chosen to analyze published articles we would not have been able to consider studies that have not been published, either because they have null findings or because of other biases inherent in the publication process (Boudreau 2021:341, Malhotra 2021:356, Jacobs 2020:240). Besides, the process of writing the PAP is an opportunity to think carefully about the design, and potentially, iterate; other considerations (such as word limits) may lead researchers to avoid providing too much detail in a manuscript.^{9 10}

To conduct our analysis, we downloaded all available PAPs submitted in 2019, including amendments and updates, and processed them using text-analysis tools. Our goal was to identify how many of these PAPs contained qualitative tools as part of the design. To do so, we searched the documents for keywords such as *fieldwork*, *archive*, *ethnography*, *interview*, *triangulation*, *oral*, *history*, *qualitative*, *participant observation*, and *process tracing*. We took the absence of such words from the design description as evidence that the project did not include a qualitative fieldwork component. The results of this search are summarized in Figure 1.

⁹ In the appendix we provide results of two sets of additional analyses: 1) a search of manuscripts based on all the PAPs that stated they used mixed-methods, plus a random sample of 10 others, for a total of 33; 2) the 2020 issue of the American Political Science Review (APSR), one of the most important journals in political science. The results of this analysis are consistent with our argument.

¹⁰ For an exception where a working paper provides further detail than the PAP, see Cooperman (2019) “Trading Favors: Local Politics and Development in Brazil.” In the appendix, she provides extensive detail about her iterative process between fieldwork and design for the refinement of theory and hypotheses in an observational study.

Figure 1: Text Analysis using Pre-Analysis Plans in EGAP, 2019.

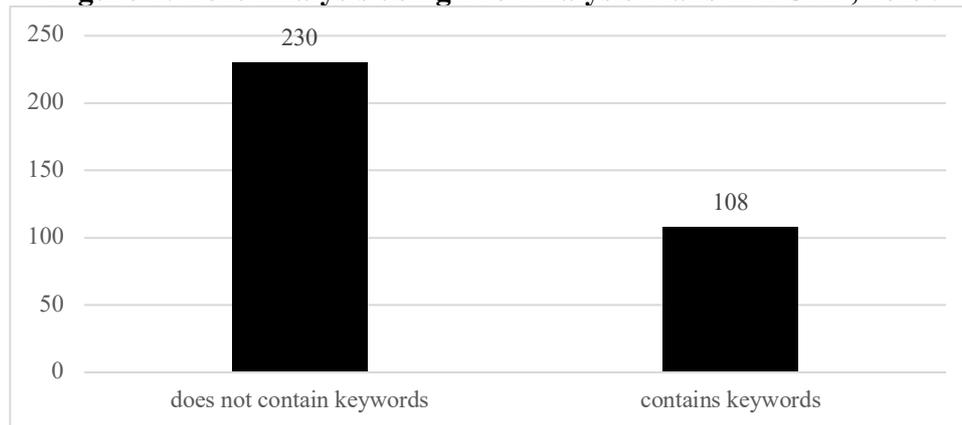
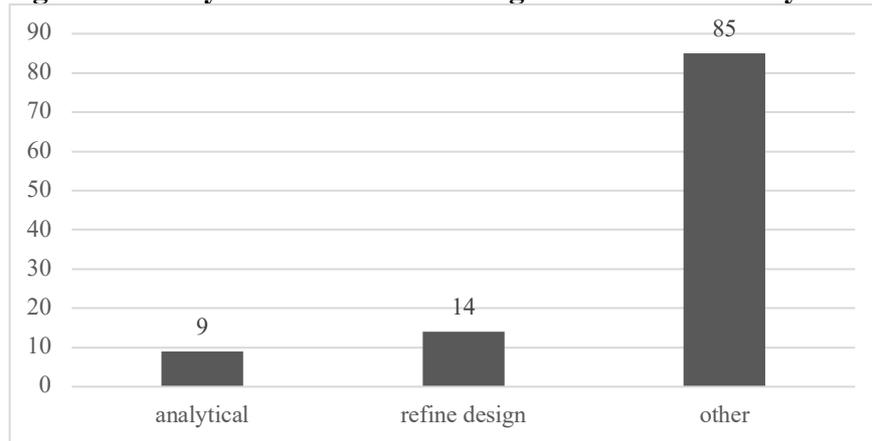


Figure 1 shows that of the 338 documents analyzed, about a third (108) contain one or more of the relevant keywords, whereas two-thirds (230) do not. We then carefully read the subset of 108 documents that contain keywords to identify each study’s purpose in combining methods. We classified the different purposes as either *analytical*, *refine design*, or *other*. “Analytical” refers to the combination of methods to identify mechanisms and build theory. “Refine design” refers to the combination of methods to improve some aspect of the research design, such as to improve the treatment, to adjust the treatment and control groups, to add information about the context, or to adjust hypotheses. The category “other” identifies those cases where terms such as “interview” refer to the administration of survey or field experiments but do not refer to interviews in a qualitative sense, as well as to documents that were misclassified; for example, the word “field” only appeared in the bibliography. The results of that additional step are presented in Figure 2.

According to Figure 2, only 14 designs combine methods for the purpose of improving the design itself, which represents only 4% of the PAPs pre-registered in EGAP in 2019. The vast majority of the designs that employ mixed methods combine interviews with experiments; other tools such as participant observation or ethnographic methods are far less common, and iteration between elements of the design is rarely discussed explicitly.

Figure 2: Analysis of PAPs containing mixed-methods keywords.



In a third and final step, we re-read the PAPs under “refine design” to identify how different methods were combined. For example, Camacho et al.’s (2019)¹¹ study of excessive alcohol consumption and violence in bars in Bogota combines a survey experiment with fieldwork in four municipalities. Insights from fieldwork were used to identify where the survey would be deployed, based on security concerns. Even though one can infer that the geographical area for survey implementation was adjusted based on qualitative fieldwork, the authors do not provide explicit details of this process. In another example, Bezzola and Brugger (2019)¹² combine interviews and focus groups with a field experiment to study the effect of mining companies’ social investment on citizens’ perceptions of local governments in two mining communities in Burkina Faso. The PAP explicitly states that fieldwork is used to inform the identification of relevant experimental outcomes and to test and improve an audio-podcast. However, the design does not specify how these steps were taken, or the modifications that were included. These examples show that researchers mix methods with the purpose of refining their research designs, and in some cases they do so iteratively. The designs include creative combinations of methods, unfortunately, the PAPs do not include

¹¹ “Excessive Alcohol Consumption and Violent Brawls: A Randomized Controlled Trial with Bartenders in Bogotá, Colombia,” registered design available here: <https://osf.io/nxh7v>

¹² “Social Investments of Mining Companies and Citizen Engagement in Local Governance,” registered design available here: <https://osf.io/8v5tw>

enough details to allow other researchers to learn from these strategies. One important caveat in our analysis is that these descriptive statistics might be biased. As we argued before, it is possible that authors leave out discussions of qualitative methods because they see them as secondary, or pre-scientific. Yet, when PAPs include such details it allows for a better understanding of the whole research process, and it gives other researchers more elements to replicate the design.

Refining the research design is essential for good causal inference. A good research design allows for good data collection and analysis; this is critical for experiments. While some observational designs can be adjusted *ex post*, this is not possible for experiments; for example, once a field experiment is implemented, researchers cannot redo it (Gerring 2011). Experimental research design can benefit from iteration with qualitative elements. Iterating allows the researcher to adjust hypotheses, as well as the treatment and measurement tools, prior to implementation. If researchers do not combine methods, they run the risk of employing unrealistic treatments. By creating realistic treatments, researchers strengthen the design's internal validity: higher correspondence between the measurement tool and the phenomenon of interest. Furthermore, when researchers combine methodologies and make the steps involved in the combination explicit, they make the design more transparent, which allows others to replicate it.

In the next section, we provide an example of a mixed-methods case study in which we refine an experimental design, as well as the subsequent analysis, in the context of a study of public opinion support for punitive policing practices in the city of Montevideo, Uruguay. We pay particular attention to iteration in the combination of qualitative fieldwork and experiments. We then discuss the contributions of the iterative process.

4. Iterating between Experiments and Fieldwork: the Case of Punitive Security Policy in Uruguay

We illustrate our proposed iterative methodology with a study of public opinion support for punitive policing practices in the city of Montevideo, Uruguay. The study combines a survey experiment with intensive fieldwork, including direct observation and interviews. The steps we describe apply more generally to other mixed-methods designs, with or without experiments.

Recently, Latin American countries have experienced a “security crisis,” associated with a rise in crime and violence. Citizens’ concern with security has pushed governments to adopt increasingly punitive responses that enjoy widespread support. Punitive policies—policing practices characterized by the escalation of the use of force, or threat thereof, and the selective application of the law—have yielded inconclusive results with respect to crime reduction (Holland 2013). Furthermore, in many cases, these policies have led to increased levels of police violence (Wolf 2017).

These seemingly divergent logics—the spread of punitive policing without positive results in terms of crime reduction—motivate the main research question in this study: why does public opinion support punitive policing? We argue that public opinion support for punitive policing is associated not with expected effectiveness of the practice, but with implicit beliefs and preferences about whether the target of punitive policing deserves such treatment. We further argue that individuals evaluate deservingness based on a target’s socio-demographic attributes.

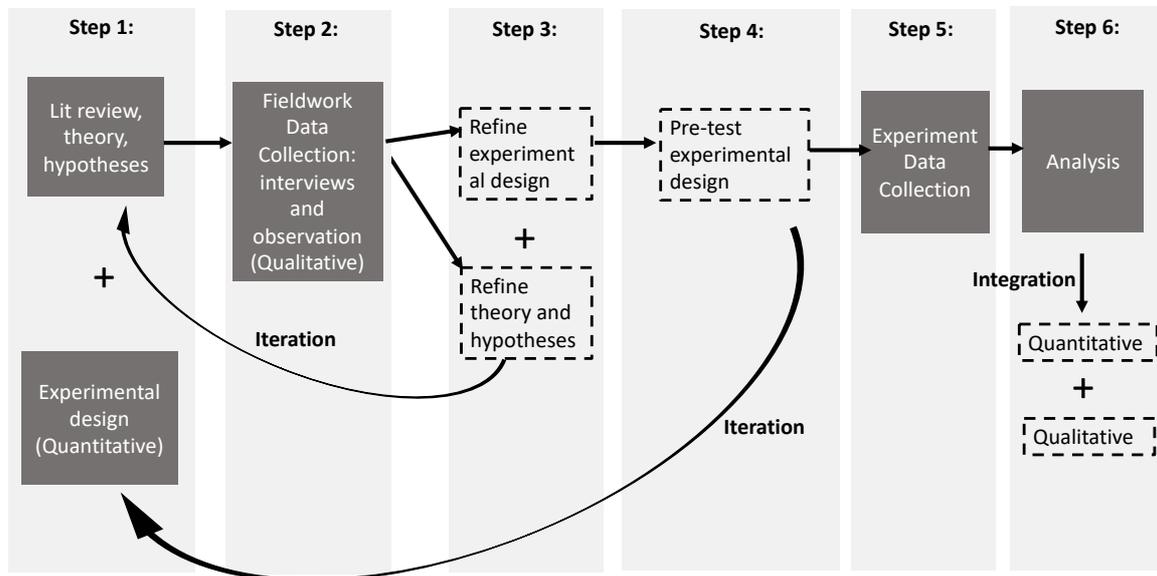
To test this argument, we conducted an online survey experiment consisting of a factorial design with attributes embedded in a vignette. We administered the survey in Montevideo, the capital city of Uruguay, in 2020. Despite the country’s relatively low level of violence, public opinion in Uruguay supports the implementation of punitive policing

practices. Even though the study of support for punitive strategies tends to focus on countries where violence is higher, such as Mexico (García-Ponce, Young, and Zeitzoff 2019), recent work has identified similar policing strategies in low violence contexts (Gingerich and Scartascini 2018). Prior to administering the survey, we iterated multiple times between theory, fieldwork, and the experimental design, which allowed us to more clearly specify the attributes of deservingness in a low violence context such as Uruguay.

The decision to implement iteration systematically followed a deductive--inductive process. Initially, we followed the mixed-methods literature stating that it is important to verify that the components of an experiment resonate with lived experiences (Seawright 2021). Following this logic, we planned to gather insights directly from interviewees to probe two sets of assumptions: 1) assumptions in the theoretical literature about the attributes of deservingness; 2) assumptions in the experimental literature about the effects of experimental manipulations. However, we came to the decision to iterate multiple times inductively, as a result of the first phase of fieldwork, where we realized that we needed to adjust the design, and then verify whether it needed further adjustments.

The following paragraphs describe the iterative process we employed in our theoretical and empirical design, emphasizing the use of qualitative fieldwork to refine the design: 1) preliminary theory and hypotheses and preliminary data collection strategies; 2) qualitative data collection (interviews, participant observation) to refine theory; 3) use of qualitative insights to refine the experimental component; 4) pre-test of quantitative component and subsequent refinement based on the results of the pre-test. We also employed qualitative data for analytical purposes (integration of qualitative and quantitative components of the design), but the iterative process takes place during the design phase (steps 1 through 4 in Figure 3), prior to testing. Figure 3 presents a summary of the entire process (design and analysis).

Figure 3: Iteration in the Refinement of Testing Tools



Preliminary theory and hypotheses and preliminary data collection strategies (step 1, no iteration)

Research designs usually start with a review of the literature and articulate a proto-theory including a set of hypotheses. Measurement instruments are selected and specified based on the variables defined in the theory.

In our project, we first developed a theory based on insights from public policy research, which suggests that citizens support certain welfare policies based on whether they believe beneficiaries deserve the benefits. Research on welfare shows that individuals base their support for social welfare benefits on their perceptions of who deserves to receive them. For example, if beneficiaries are perceived as “lazy,” individual support for welfare policies decreases. Deservingness is a cognitive shortcut; it operates by reinforcing pre-conceived notions based on incomplete information (Petersen 2012).

We argue that much like welfare policy, support for punitive policing policies is based on perceptions of deservingness. Individuals have preconceived ideas about who is likely to be a criminal. Citizens see certain groups as criminal and as less deserving of guarantees of due process. Because deservingness is an abstract, complex concept, individuals use sociodemographic cues (as attributes of deservingness) to determine who does not deserve due process. For example, due to racial stereotypes, a black individual will be perceived as dangerous, likely to be a criminal, and thus, as less deserving of due process. These attributes are contextually constructed; therefore, they may vary across contexts.

To identify which attributes of deservingness are relevant in a given context, we initially developed a forced-choice conjoint experiment. Conjoint experiments are popular in political science and they are frequently used to disentangle multidimensional preferences. They also minimize social desirability bias in survey responses. In regular surveys, respondents may be reluctant to state how they truly feel about individual features and police actions. Embedding each characteristic in a larger set of features and assigning them randomly to subjects may help reduce social desirability bias. Forced-choice conjoint experiments are usually embedded in a survey and consist of presenting respondents with a list of individual attributes and asking them to make a hypothetical choice (Hainmueller, Hopkins, and Yamamoto 2014). As explained in more detail below, after pre-testing our original design, we ended up presenting respondents with a vignette describing a hypothetical interaction between police and an individual in our final design. We also included an image of the purported individual, randomizing socio-demographic and appearance features. We then asked respondents to evaluate police behavior.

Qualitative Data Collection (step 2)

Strategies for data collection can have multiple purposes in relation to the research design. In many cases, researchers collect data to test a theory. Qualitative data can also be used to build a theory, provide additional contextual information, or to refine hypotheses (Yom 2015).

We developed our argument in several iterations. After specifying our theory deductively, we used information obtained during fieldwork and we went back to refine the theory; this constitutes a first iteration in our design. We conducted our fieldwork over a period of three months in a pre-selected set of neighborhoods of Montevideo. We conducted interviews with about 40 individuals: government officials, police officers, community leaders, representatives of local governments, and ordinary neighbors. The purpose of the interviews was to obtain evidence about how the deservingness mechanism operates.

As a result of the interviews, we were better able to understand respondents' perceptions of police behavior as well as their ideas about who deserves to be treated differently by the police. We also attended community meetings and reviewed documentation concerning security policy implementation. In addition, we reviewed press articles from four different newspapers to better grasp the context of crime, insecurity and police responses in Montevideo. These different sources were useful for triangulating and verifying information. We used insights from the interviews to refine our hypotheses and adjust the theory, and to refine the experimental design, as described below.

Refinement of Theory and Experimental Component (step 3, iteration)

In our original design, our intuition was that deservingness could be tied to attributes such as nationality, race, or location. However, because the concrete manifestation of some of these cleavages was not clear, and some of them are not particularly salient in the Uruguayan context (neither race nor xenophobic cleavages are particularly strong), we needed to use

fieldwork systematically to refine our theory by identifying the relevant attributes of deservingness. Because deservingness is abstract and can be understood differently in different contexts, we needed to ensure that we were gathering evidence about it, and not some other related phenomenon. In many of our interviews, neighborhood residents stated that young, dark-skinned men who happened to be wearing baseball caps or hoodies, were a symbol of insecurity. For example:

“I say to [my adolescent son] ‘If you came up to me, and you weren't my son, I'd think you were going to rob me.’ He says, ‘But I don't dress bad!’ ‘No, but you wear that hat, those jeans, that jacket, and when you dress like that you look just like the rest of them.’”¹³

Based on these insights, we adjusted the theory to incorporate social class as a relevant attribute of deservingness. We take attire to serve as a proxy for social class (Wolf 2017). This insight also led us to incorporate “attire” as a marker of deservingness in the set of descriptive features used in the experiment, which we would not have done had we not conducted interviews and informal conversations with neighborhood residents.

Qualitative data in mixed-methods designs can also be used to refine the quantitative tools. For treatments to have an effect they should first resonate with individuals, they should be realistic (Seawright 2021). To improve our experiment, we incorporated observations from fieldwork by, 1) identifying relevant attributes to test in the experimental design; and, 2) incorporating descriptions of police actions and interactions with individuals to create more realistic scenarios.

Here, we iterated again. In the original design, we selected a set of seven socio-demographic features—based on our review of the literature—as attributes associated with deservingness: sex, age, nationality, education, race, religion, and whether the individual had

¹³ Interview, neighborhood resident, June 27, 2019.

a criminal record. We also included actions by the hypothetical individual and police, as well as the neighborhood in which the encounter took place (see the appendix for a full description of the design). Once in the field, we used interviews and observations to validate the design. To ensure we had identified relevant attributes and situations, we asked interviewees to describe the kinds of interactions they had with the police (their own, or ones they knew about or had witnessed), and to provide information of the context in which the interactions occurred and of those involved. As a result, we learned about relevant manifestations of deservingness and common police behaviors. We then revised the characteristics listed in the experiment based on those descriptions, for example:

“they are quite violent, they stopped a bus, got on (...) they wanted to arrest an adolescent who was on the bus with his uncle because they did not like the way he looked at them, they ended up throwing punches, the whole bus witnessed it”¹⁴

“they would stop you on your way to work, take your backpack and throw all its contents on the floor, they were very aggressive”¹⁵

We adjusted the original design based on the information from interviews. By contextualizing our design, we ensured that the instrument is an adequate measure of our concepts of interest (Steiner, Atzmüller, and Su 2016). We employed an experiment, but mixed methods can be used to combine indicators such as scales, for example, or to identify new variables to include in regression analysis.

Pre-tests (step 4, iteration)

Any research design containing a questionnaire benefits from pre-testing to check the mechanics and to ensure that the questions capture the phenomena of interest. Before

¹⁴ Interview, government official, August 1, 2019.

¹⁵ Interview, neighborhood resident, June 13, 2019.

administering the survey, we conducted a pre-test: we shared the survey with a small group to test its mechanics, and to evaluate our experimental design. We gathered data from about 35 respondents (undergraduate students and faculty members) and we asked them to take the survey and comment on the format and on anything they found unusual.

The typical presentation of a forced-choice conjoint experiment includes two profiles that are being compared on a list of attributes. In our case, we presented respondents with two hypothetical individual profiles. Each profile comprised a series of socio-demographic attributes that varied at random, as well as a series of actions the police and a hypothetical individual were engaged in. After reviewing the comments, we concluded that this format made it hard for respondents to understand what they were expected to do, as several respondents mentioned they found the question hard to follow and a bit confusing. For example, two respondents stated that it was hard for them to compare the profiles, another one mentioned that she did not quite understand the sequence of actions, and yet another one suggested that a photo would summarize the physical characteristics more easily.

Based on these results, we went back to the original design, and switched it: instead of presenting respondents with a list of attributes, we presented a vignette with an image. By substituting an image for the list of attributes, we reduced some of the task complexity because an individual's attributes were summarized in a photo. This change reduced the number of dimensions that respondents needed to compare and made it more engaging and easier to understand. Instead of asking respondents to compare between profiles, we presented the vignette four times, varying the attributes. We present a comparison between the two designs in the appendix.

To obtain the photos, we hired actors. Once we adjusted the design incorporating the images, we conducted a manipulation check as part of the pre-test. Scholarly work on experimental research emphasizes the importance of conducting “manipulation checks,” to

ensure that the experimental treatment is working as expected (Mutz 2021). The manipulation check consisted of asking respondents in the pre-test to rate the images based on a series of characteristics (race, dangerousness, social class, attractiveness) to ensure that our treatments of interest were actually having an effect, and to rule out confounding factors. We showed respondents 15 images of 5 actors. Based on the manipulation check, we selected the specific images we presented. We finally pre-tested the whole survey again with the new experimental design.

Quantitative Data Collection and Analysis (step 5 and 6, analytical)

The last step is the implementation of the quantitative component and subsequent data analysis. In this stage, qualitative insights can be used to amplify the quantitative results. We administered our survey online between January 29 and March 17, 2020, targeting individuals at least 18 years old who resided in Montevideo. We collected about 2,900 responses. In the analytical phase of the research process, rather than iterating, we combined methodological approaches through the process of integration.

We integrated the analysis of the experimental results with insights from our interviews. The experiment produced two important findings. First, whether someone is seen as deserving more violent treatment from the police depends on the individual's social class and their actions. Second, regardless of the identity and behavior of the individual in the vignette, respondents justify stop-and-frisk as a reasonable police action. Because the experiment does not tell us why individuals see this police practice as reasonable, we drew insights from our interviews to better contextualize this finding.

During our conversations with police officers and neighborhood residents, we learned that stop-and-search is a very common practice. People are used to seeing the police conduct this activity, which suggests that this finding might be the result of a normalized routine

practice. Residents of neighborhoods in Montevideo also state that police stop individuals based on the way they look—i.e. “profiling”—a practice they deem necessary to combat crime.

This study could have been conducted by constructing the experiment based only on prior research, without incorporating fieldwork. However, successive iterations between fieldwork and theory, and between fieldwork and design and analysis, resulted in a much richer research design and analysis, and stronger causal inference. Fieldwork revealed the relevance of attributes we had not considered, and the irrelevance of others we previously thought were important. Though we have described our research process as a linear sequence of steps, the iterative process is far from linear, as captured in Figure 3.

5. Extending Iteration Beyond Experimental Designs

In our study of attitudes towards punitive policing, we improved the operationalization and measurement of our variables through iteration. Insights from fieldwork allowed us to identify and measure attributes of the concept of deservingness. Using interviews, we identified the most relevant attributes and adapted the experiment accordingly. Furthermore, because the notion of deservingness varies by context (in some cases it may be more closely linked to race while in other cases it may be linked more to social class, or to other attributes), using only deduction to select which attributes to study may lead to erroneous measurement and interpretation. The refinement of testing tools lies somewhere in between theory building and testing, and it involves more than pre-testing of instruments for data collection.

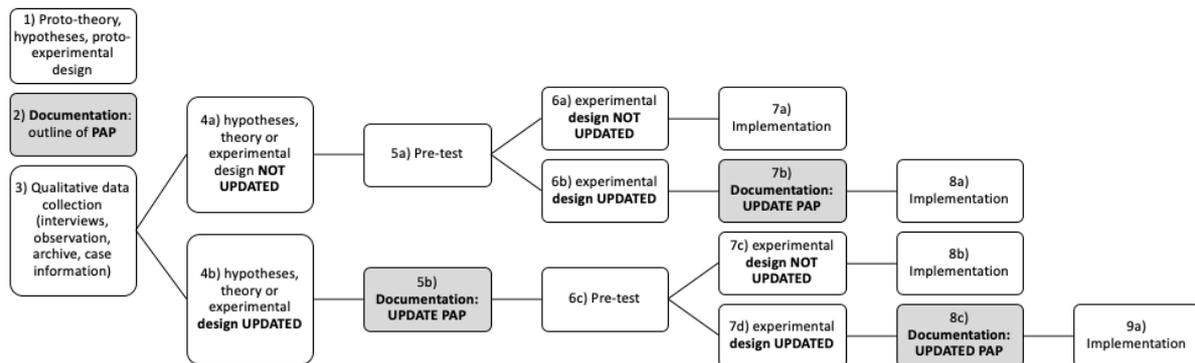
Making iteration explicit in an experimental context allows for increased transparency in research design because researchers can show all the steps involved in the construction of

the design. This process facilitates replication of research in different contexts, as it makes clear to researchers which parts of the design should be adapted.

Although researchers value the benefits of iteration, the way in which it occurs is rarely made explicit, which obscures its benefits. Iteration also calls attention to research design as a non-linear process: it requires alternating back-and-forth between different elements of the design, where researchers update prior information (Dunning 2008; Kapiszewski, MacLean, and Read 2015; Yom 2015). When an experiment includes insights from fieldwork, the researcher ensures that experimental results are not an artifact of the research design, thus lending added credibility to the findings.

To make iteration explicit, researchers should document each step in the process and pre-register the entire process in a PAP. Figure 4 demonstrates a series of practical steps researchers can take to ensure proper iteration and documentation when combining fieldwork and experimental designs. The key issue is to ask whether theory, hypotheses and/or design need to be updated in light of fieldwork. If any of these elements need to be updated, another documentation phase is required, the previous PAP needs to be revised. In revising the PAP, researchers should explicitly state what evidence leads to the proposed change. If none of these elements are updated, the research process continues through a pre-test. In light of the pre-test, a similar question is asked: does the experimental design need to be updated? The researcher should repeat the same steps, documenting as needed, until quantitative data collection can begin. Documentation and updating of PAPs is fundamental to ensure transparency and replicability of the research process as a whole.

Figure 4: Steps to Implement Iterative Designs with Fieldwork and Experiments



Our approach is applicable to designs with other types of quantitative elements.

Iteration works to strengthen theory and data collection tools and can be applied to designs containing interviews or survey questionnaires which are then used in, for example, regression analyses. The choice of qualitative technique to inform quantitative methods is at the service of the research question. In our example, the aim was to understand the attributes of deservingness to improve measurement, thus we needed to speak to individuals. We conducted one- on- one and group interviews to refine our theory. However, we could have taken an inductive starting point and used ethnographic methods instead. Thachil (2018) and Dunning and Harrison (2010) used qualitative methods to build their respective samples. Thachil uses ethnography to identify the relevant socio-demographic attributes of his sample. The specific method may vary based on what tools the researcher needs to answer her research question and to probe the assumptions behind other methods (Seawright 2016). Yet, these choices may also be determined by pragmatic considerations: time, budget and access. Researchers should also factor in these elements when deciding how to iterate.

6. Conclusions

This paper highlights the value of iteration between different elements of the research design to improve research combining qualitative elements and experiments, prior to testing.

We seek to fill a gap in the literature, namely, qualitative research that highlights the value of iteration does not address experimental work, and, in general, experimental work does not combine methods explicitly. According to our analysis of EGAP's pre-analysis plans, most experiments do not make the process of combining methods explicit. However, researchers would benefit from making the process more transparent.

We make three contributions: First, we highlight the role of iteration in improving causal inference in mixed-methods research designs using experiments. Related, the project demonstrates improvements to concept operationalization and measurement gained by paying attention to the context. Second, we showcase the importance of making explicit the incorporation of qualitative elements in experimental designs. By outlining the steps involved in iteration, we also contribute to discussions of transparency in research design. Third, the paper offers insights to better understand the conditions under which experimental work is generalizable. More generally, the steps we propose can apply to other designs that combine qualitative and quantitative methods.

Iteration helps a researcher to improve causal inference. For example, researchers usually begin the research process by proposing a theory, a few hypotheses, and a research design to test the argument. The researcher can choose to take a linear path between theory and testing, or to iterate. If she chooses the latter, she can, for example, use interviews to better understand the context of the experimental intervention. A better understanding of the context allows her to refine her hypotheses and create more realistic interventions, thus strengthening internal validity. Iteration is also relevant in non-experimental designs, where researchers start with deduction and refine their hypotheses in light of their interaction with data (Yom 2015).

Field research remains underutilized as a tool to improve causal inference and other aspects of the research design. This may result from a poor understanding of its advantages:

aside from strengthening causal inference, iterating with fieldwork produces more realistic treatments because it increases contextual knowledge. In addition, being explicit about iteration makes research more transparent and more easily replicated. By documenting the steps involved in iteration and making the documentation public in a PAP allows for other researchers to replicate the design and to learn from other researchers' process regardless of publication status.

One of our project's key contributions concerns the ability to generalize in experimental work. As mentioned before, one of the limitations of experimental designs is the context dependency of treatment effects. However, experimental researchers are concerned about assessing the extent to which experimental designs can "travel." Along these lines, Blair and McClendon (2021) identify a series of shortcomings in current approaches for dealing with generalizability in experimental contexts. They propose a design-driven approach, using multi-arm bandit algorithms, to evaluate the generalizability of treatment effects. We suggest that iteration with fieldwork can also help ameliorate one of the most important limitations the authors highlight, namely how treatments are assigned to contexts.

Blair and McClendon (2021: 421) explain that researchers usually use "intuition" to choose the most effective intervention for a given context. This is problematic because, aside from eliminating other alternative treatments, it implies that pre-conceived notions about what might work are the basis for the design. Our work, however, shows that conducting fieldwork allows researchers to design treatments that take the context into consideration, thus minimizing the incorporation of researchers' pre-conceived notions in the research design. As Robert Bates has noted, conducting fieldwork allows researchers to change preconceived notions they might have and to adapt their theories to reality (Munck and Snyder 2007:25). A good understanding of contextual features helps clarify the scope of the theory being tested, and thus, where else it might apply.

Even though iterating between fieldwork and experiments has many advantages, there are some practical limitations: costs, knowing when to stop, and the possibility of inducing confirmation bias. Fieldwork is time-consuming and costly, it entails coordinating meetings, interviews, traveling to remote or difficult-to-access locations, and may even jeopardize researchers' safety.¹⁶ However, even under these circumstances, iteration is still possible. Other qualitative tools can be incorporated to understand the context and improve designs. For example, researchers can use information from newspapers or secondary literature about cases to contextualize their experimental interventions. Another risk that researchers should be aware of is the tendency to continue to refine the design indefinitely. A good place to stop is when field observations no longer provide new evidence to refine the design. For example, when information from interviews or observation begins to be repeated across different respondents (Saunders et al 2018). Last, researchers may wonder about whether incorporating realism might introduce demand effects—where respondents may guess what the expected response should be and act accordingly (Mummolo and Peterson 2019). We believe that concerns about priming participants are mitigated when researchers employ qualitative methods to enhance experimental realism. No method is completely exempt from confirmatory bias, but research suggests that the risk of generating demand effects in survey experiments is generally low (Mummolo and Peterson 2019). In addition, presenting real-life situations based on qualitative evidence instead of manufactured ones mitigates priming biases by accurately representing the context (Carpenter, Montgomery and Nylen 2020).

Iteration in mixed-methods designs can happen in multiple ways, and it has slowly become more widely used. Researchers concerned with improving causal inference and

¹⁶ This challenge may be more acute for women researchers, who may face a higher risk of experiencing violence than men.

producing transparent, replicable designs, would benefit from explicitly including iteration as part of their research process.

7. References

- Auerbach, Adam, and Tariq Thachil. 2020. "Cultivating Clients: Reputation, Responsiveness, and Ethnic Indifference in India's Slums." *American Journal of Political Science* 64(3): 471-487.
- Beach, Derek, and Levente Littvay. 2020. "A More Conscious Experimental Design: How More Explicit Use of Process Tracing Can Improve Experiments", 116th APSA Annual Meeting, September 30 – October 3.
- Blair, Graeme, Jasper Cooper, Alexander Coppock, and Macartan Humphreys. 2019. "Declaring and Diagnosing Research Designs." *American Political Science Review* 113 (3): 838–59. <https://doi.org/10.1017/S0003055419000194>.
- Blair, Graeme, Alexander Coppock and Macartan Humphreys. Forthcoming. *Research design: Declaration, diagnosis, redesign*. Princeton University Press.
- Blair, Graeme, and Gwyneth McClendon. 2021. "Conducting Experiments in Multiple Contexts." In *Advances in Experimental Political Science.*, edited by James N. Druckman and Donald P. Green, 411–428. New York: Cambridge University Press.
- Boudreau, Cheryl. 2021. "Transparency in Experimental Research." In *Advances in Experimental Political Science.*, edited by James N. Druckman and Donald P. Green, 339–353. New York: Cambridge University Press.
- Brady, Henry, and David Collier. 2010. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Maryland: Rowman & Littlefield Publishers.
- Carpenter, Charli, Alexander H. Montgomery, and Alexandria Nylen. 2021. "Breaking Bad? How Survey Experiments Prime Americans for War Crimes." *Perspectives on Politics* 19 (3): 912–24. <https://doi.org/10.1017/S1537592720002376>.
- Chappuis, Fairlie and Jana Krause. 2019. "Research Dilemmas in Dangerous Places". In De Goede, Marieke, Esmé Bosma, and Polly Pallister-Wilkins, eds. *Secrecy and*

- Methods in Security Research: A Guide to Qualitative Fieldwork*. Routledge, 2019: 112-125.
- Clayton, Amanda, Amanda Lea Robinson, Martha C. Johnson, and Ragnhild Muriaas. 2020. “(How) Do Voters Discriminate Against Women Candidates? Experimental and Qualitative Evidence From Malawi.” *Comparative Political Studies* 53 (3–4): 601–30. <https://doi.org/10.1177/0010414019858960>.
- Cyr, Jennifer. 2019. An integrative approach to measurement: focus groups as a survey pretest. *Qual Quant* 53, 897–913. <https://doi.org/10.1007/s11135-018-0795-5>
- Deaton, Angus, and Nancy Cartwright. 2018. “Understanding and Misunderstanding Randomized Controlled Trials.” *Social Science & Medicine, Randomized Controlled Trials and Evidence-based Policy: A Multidisciplinary Dialogue*, 210 (August): 2–21. <https://doi.org/10.1016/j.socscimed.2017.12.005>.
- Druckman, James N., and Donald P. Green. 2021. “A New Era of Experimental Political Science.” In *Advances in Experimental Political Science.*, edited by James N. Druckman and Donald P. Green, 1–15. New York: Cambridge University Press.
- Dunning, Thad. 2008. “Natural and Field Experiments: The Role of Qualitative Methods.” *Qualitative & Multi-Method Research* Fall: 17–23.
- . 2012. *Natural Experiments in the Social Sciences: A Design-Based Approach*. New York: Cambridge University Press.
- . 2015. “Improving Process Tracing: The Case of Multi-Method Research.” In *Process Tracing. From Metaphor to Analytic Tool*, edited by Andrew Bennett and Jeffrey T. Checkel, 211–36. Nueva York: Cambridge University Press.
- Dunning, Thad, and Lauren Harrison. 2010. “Cross-Cutting Cleavages and Ethnic Voting: An Experimental Study of Cousinage in Mali.” *The American Political Science Review* 104 (1): 21–39. <https://doi.org/10.2307/27798538>.

- Elman, Colin, John Gerring, and James Mahoney. 2020. *The Production of Knowledge: Enhancing Progress in Social Science*. Cambridge University Press.
- Fairfield, Tasha, and Andrew E. Charman. 2017. "Explicit Bayesian Analysis for Process Tracing: Guidelines, Opportunities, and Caveats." *Political Analysis* 25 (3): 363–80. <https://doi.org/10.1017/pan.2017.14>.
- Ferwerda, Jeremy, and Nicholas L. Miller. 2014. "Political Devolution and Resistance to Foreign Rule: A Natural Experiment." *American Political Science Review* 108 (3): 642–60. <https://doi.org/10.1017/S0003055414000240>.
- García-Ponce, Omar, Lauren Young, and Thomas Zeitzoff. 2019. "Anger and Support for Punitive Justice in Mexico's Drug War." APSA Annual Meeting, Washington DC.
- Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W. W. Norton.
- Gerring, John. 2011. *Social Science Methodology: A Unified Framework*. New York: Cambridge University Press.
- Gingerich, Daniel, and Carlos Scartascini. 2018. "A Heavy Hand or a Helping Hand? Information Provision and Citizen Preferences for Anti-Crime Policies." Inter-American Development Bank.
- Goertz, Gary. 2017. *Multimethod Research, Causal Mechanisms, and Case Studies: An Integrated Approach*. New Jersey: Princeton University Press.
- Gonzalez-Ocantos, Ezequiel, and Jody LaPorte. 2019. "Process Tracing and the Problem of Missing Data." *Sociological Methods & Research*, February. <https://doi.org/10.1177/0049124119826153>.
- Gonzalez-Ocantos, Ezequiel and Juan Masullo J. 2020. "Aligning Interviewing with Process Tracing", working paper.
- Hainmueller, Jens, Daniel J. Hopkins, and Teppei Yamamoto. 2014. "Causal Inference in

- Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments.” *Political Analysis* 22 (1): 1–30. <https://doi.org/10.1093/pan/mpt024>.
- Harbers, Imke, and Matthew C Ingram. 2020. “Mixed-Methods Designs.” In *The SAGE Handbook of Research Methods in Political Science and International Relations*, edited by Luigi Curini and Robert Franzese. United Kingdom: SAGE.
- Holland, Alisha C. 2013. “Right on Crime?: Conservative Party Politics and Mano Dura Policies in El Salvador.” *Latin American Research Review* 48 (1): 44–67. <https://doi.org/10.1353/lar.2013.0009>.
- Hollstein, Betina. 2014. “Mixed Methods Social Networks Research: An Introduction.” In *Mixed Methods Social Networks Research. Design and Applications*, edited by Betina Hollstein and Silvia Domínguez. New York: Cambridge University Press.
- Jackson, Michelle, and David R. Cox. 2013. "The principles of experimental design and their application in sociology." *Annual Review of Sociology* 39: 27-49.
- Jacobs, Alan M. 2020. “Pre-Registration and Results-Free Review in Observational and Qualitative Research.” In *The Production of Knowledge. Enhancing Progress in Social Science*, edited by Colin Elman, John Gerring, and James Mahoney. New York: Cambridge University Press.
- Jha, Saumitra, Vijayendra Rao, and Michael Woolcock. 2007. “Governance in the Gullies: Democratic Responsiveness and Leadership in Delhi’s Slums.” *World Development, Experiences of Combining Qualitative and Quantitative Approaches in Poverty Analysis*, 35 (2): 230–46. <https://doi.org/10.1016/j.worlddev.2005.10.018>.
- Kapiszewski, Diana, Lauren M. MacLean, and Benjamin L. Read. 2015. *Field Research in Political Science*. Cambridge: Cambridge University Press.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton University Press.

- Kocher, Matthew A., and Nuno P. Monteiro. 2016. "Lines of Demarcation: Causation, Design-Based Inference, and Historical Research." *Perspectives on Politics* 14 (4): 952–75. <https://doi.org/10.1017/S1537592716002863>.
- Levy Paluck, Elizabeth. 2010. "The Promising Integration of Qualitative Methods and Field Experiments." *The ANNALS of the American Academy of Political and Social Science* 628 (1): 59–71. <https://doi.org/10.1177/0002716209351510>.
- Lieberman, Evan S. 2005. "Nested Analysis as a Mixed-Method Strategy for Comparative Research." *American Political Science Review* 99 (3): 435–52. <https://doi.org/10.1017/S0003055405051762>.
- Lupia, Arthur, and Colin Elman. 2014. "Openness in Political Science: Data Access and Research Transparency: Introduction." *PS: Political Science & Politics* 47 (1): 19–42. <https://doi.org/10.1017/S1049096513001716>.
- Malhotra, Neil. 2021. "Threats to the Scientific Credibility of Experiments: Publications Bias and P-Hacking." In *Advances in Experimental Political Science.*, edited by James N. Druckman and Donald P. Green, 354–368. New York: Cambridge University Press.
- Mummolo, Jonathan, and Erik Peterson. 2019. "Demand Effects in Survey Experiments: An Empirical Assessment." *American Political Science Review* 113 (2): 517–29. <https://doi.org/10.1017/S0003055418000837>.
- Munck, Gerardo L., and Richard Snyder, eds. 2007. *Passion, Craft, and Method in Comparative Politics*. JHU Press.
- Mutz, Diana. 2021. "Improving Experimental Treatments in Political Science." In *Advances in Experimental Political Science.*, edited by James N. Druckman and Donald P. Green, 219–238. New York: Cambridge University Press.
- Pérez Bentancur, Verónica, Rafael Piñeiro Rodríguez, and Fernando Rosenblatt. 2018. "Unexplored Advantages of DART for Qualitative Research." *Qualitative & Multi-*

- Method Research* 16 (2): 31–35. <https://doi.org/10.5281/zenodo.3524354>.
- Petersen, Michael Bang. 2012. "Social Welfare as Small-Scale Help: Evolutionary Psychology and the Deservingness Heuristic." *American Journal of Political Science* 56 (1): 1–16. <https://doi.org/10.1111/j.1540-5907.2011.00545.x>.
- Redlich Revkin, Mara. 2020. "What Explains Taxation by Resource-Rich Rebels? Evidence from the Islamic State in Syria." *Journal of Politics* forthcoming.
<https://doi.org/10.2139/ssrn.3023317>.
- Saunders, Benjamin, Julius Sim, Tom Kingstone, Shula Baker, Jackie Waterfield, Bernadette Bartlam, Heather Burroughs, and Clare Jinks. 2018. "Saturation in qualitative research: exploring its conceptualization and operationalization." *Quality & Quantity*, 52(4), 1893-1907.
- Seawright, Jason. 2021. "What Can Multi-Method Research Add to Experiments?" In *Advances in Experimental Political Science.*, edited by James N. Druckman and Donald P. Green, 369–384. New York: Cambridge University Press.
- Seawright, Jason. 2016. *Multi-Method Social Science: Combining Qualitative and Quantitative Tools*. New York: Cambridge University Press.
- Seawright, Jason, and John Gerring. 2008. "Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options." *Political Research Quarterly* 61 (2): 294–308. <https://doi.org/10.1177/1065912907313077>.
- Steiner, Peter M., Christiane Atzmüller, and Dan Su. 2016. "Designing Valid and Reliable Vignette Experiments for Survey Research: A Case Study on the Fair Gender Income Gap." *Journal of Methods and Measurement in the Social Sciences* 7 (2): 52–94. <https://doi.org/10.2458/v7i2.20321>.
- Thachil, Tariq. 2018. "Improving surveys through ethnography: Insights from India's urban periphery." *Studies in Comparative International Development* 53.3: 281-299.

- Rao, Vijayendra, Kripa Ananthpur, and Kabir Malik. 2017. "The Anatomy of Failure: An Ethnography of a Randomized Trial to Deepen Democracy in Rural India." *World Development* 99 (November): 481–97
<https://doi.org/10.1016/j.worlddev.2017.05.037>.
- Wolf, Sonja. 2017. *Mano Dura: The Politics of Gang Control in El Salvador*. Austin: University of Texas Press.
- Yom, Sean. 2015. "From Methodology to Practice: Inductive Iteration in Comparative Research." *Comparative Political Studies* 48 (5): 616–44.
<https://doi.org/10.1177/0010414014554685>.

Appendix

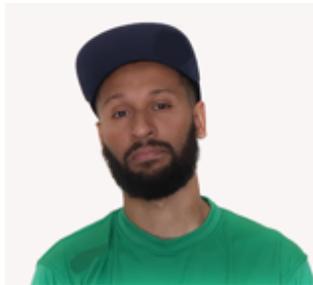
A. Original Survey Design

Individual A	Attributes	Individual B
black	<i>race</i>	white
no	<i>criminal record</i>	yes
	<i>age</i>	
male	<i>sex</i>	female
no religion	<i>religion</i>	catholic
Uruguay	<i>country of birth</i>	Dominican Republic
highschool	<i>education</i>	some school
Casavalle	<i>neighborhood where encounter happened</i>	Pocitos
physical violence attacks police	<i>police action individual action</i>	shoot walks

In which case is police action justifiable?

To what extent is police action justifiable in the case of Individual A/B?

B. Survey Design Implemented



A police car was parked in a corner, in a [middle/upper/lower] class neighborhood. The person in the photo [**walked by the police car/walked by the police car and insulted the officers in the car/was painting graffiti on a store wall/ran past the police car after having robbed a store**]. Two officers got out of the car and they [**stopped the individual and asked for his documents/stopped the individual, search his belongings and toss them on the ground/stopped the individual and gave him a beating/shot at the individual**]

On a scale of 1 to 7, where 1 is very unreasonable, and 7 is very reasonable, how do you evaluate police action given these circumstances?

C. Replication Code for Text Analysis

We utilized 338 pre-analysis plans (PAPs) registered with EGAP in 2019, and available at the time of writing the article, as input for conducting text analysis. The R code utilized to conduct the text analysis is below. The text analysis produces a first classification of the documents and identifies our relevant keywords. Those keywords are then exported to an excel file. The excel file also contains the PAP IDs which identify the documents we downloaded from EGAP's website. Any given document can contain more than one keyword, in a subsequent step, we deleted the duplicate documents in the excel file and produced Figures 1 and 2 with the remaining data.

```
library(tokenizers)
library(dplyr)
library(readr)
library(qdap)
library(tm)
library(pdftools)
library(pdfsearch)
library(tidytext)

# Read PDF documents
files <- list.files(pattern = "pdf$")
pap <- lapply(files, pdf_text)
length(pap)
corp<- Corpus(URISource(files),
              readerControl = list(reader = readPDF))

corp <- tm_map(corp, removePunctuation, ucp = TRUE)
corp<- tm_map(corp, removeWords, c("will", "brazilian", "regina", "nonblack",
"costbenefit", "nonprofit",
"nelli", "nikhar", "salt", "beat"))

pap.tdm <- TermDocumentMatrix(corp,
                              control =
                                list(stopwords = TRUE,
                                     tolower = TRUE,
                                     stemming = TRUE,
                                     removeNumbers = TRUE,
                                     bounds = list(global = c(3, Inf))))

inspect(pap.tdm[1:10,])

# Function to search for individual keywords, allows us to do a quick search for specific key
words in the whole corpus of text data from the 338 pdf documents.

my_func <- function(data, word){
  slam::row_sums(data[data$dimnames$Terms == word, ])
}
```

```

my_func(pap.tdm, "experiment")
my_func(pap.tdm, "survey")
my_func(pap.tdm, "qualitative")
my_func(pap.tdm, "treatment")
my_func(pap.tdm, "histori")
my_func(pap.tdm, "mixed")
my_func(pap.tdm, "interview")
my_func(pap.tdm, "fieldwork")
my_func(pap.tdm, "ethnograph")
my_func(pap.tdm, "integrat")
my_func(pap.tdm, "iterat")
my_func(pap.tdm, "triangulat")
my_func(pap.tdm, "participant")
my_func(pap.tdm, "archiv")
my_func(pap.tdm, "oral")

# Transform to tidy text

pap.tidy <-tidy(pap.tdm)

pap.count <- pap.tidy %>%
  count(document, term, sort=TRUE)

# Identify whether a given pdf document contains a keyword of interest
pdfs_method <- pap.count %>%
  mutate(multimethod = ifelse(term %in% c("fieldwork", "archiv", "ethnograph", "interview",
    "triangulat", "participant", "oral", "hist", "qualitat"), "yes", "no"))

# Export to Excel table
write.csv(pdf_method, "methods.csv", row.names = FALSE)

```

D. Additional Analyses:

It is possible that PAPs may not contain sufficient information about how the qualitative elements were incorporated in the designs. To further explore this possibility, we gathered additional information about the designs that were classified as mixed-method in our original analyses. In addition, to explore the possibility that some of the other PAPs could contain mixed-methods designs without stating so, we randomly selected 10 additional designs, for a total of 33 PAPs from EGAP. Another possible concern is that our PAPs might be too recent. To tackle this possibility, we also reviewed all papers in all issues published in the American Political Science Review (APSR) in 2020, a total of 75, of which we identified 13 that used experiments.

We began by searching for published papers or working papers corresponding to the 33 PAPs from EGAP. This search resulted in 14 papers, of which 5 used qualitative methods to refine the research design, 6 did so for analytical purposes, and 3 papers did not include multi-methods designs, they only contained an experiment. Table X summarizes our findings.

paper	refine design	analytical	other
yes	5	6	3
no	9	3	7

All but one paper containing multi-methods designs provided similar descriptions of the designs to those contained in the PAP, without including additional information regarding the role of qualitative elements in refining the design. The exception is the working paper “Trading Favors: Local Politics and Development in Brazil” (Alicia Cooperman, 2019).¹⁷ The author provides extensive details about the iterative nature of her research design in the paper about how qualitative fieldwork was used to improve the theory, hypotheses and experimental design (see Appendix, pp. 2-11).

To complement this analysis, we evaluated 75 papers in the 2020 issue of the American Political Science Review (APSR), of which 13 contain experimental designs. Out of those 13 papers, only 2 incorporated qualitative elements in the design in order to refine it. In this case, we reverse engineered our analysis and looked for the PAPs, to compare them to the published papers. We found only one PAP. When comparing the PAP and the paper we found that the paper contained additional details that were not included in the PAP about the incorporation of qualitative elements (Lyall, Zhou and Imai, 2020, pp. 130-131)¹⁸. However, this information would not be sufficient to adequately replicate the analysis.

These additional analyses suggest that experimental research incorporates qualitative elements in a small proportion, at least explicitly, and it is not detailed enough to allow for proper replication of the steps taken to refine the design.

¹⁷ Link to working paper in author’s web page: https://988584bf-d53d-4c3b-ac5c-b2851e8bfb5c.filesusr.com/ugd/588c1b_32abf0b1e120416497bc309958a7203c.pdf

¹⁸ Jason Lyall, Yang Yang Zhou, Kosuke Imai, “Can Economic Assistance Shape Combatant Support in Wartime? Experimental Evidence from Afghanistan,” APSR 114 (1): 126-143.