FGV EBAPE

RAP BRAZILIAN JOURNAL OF PUBLIC ADMINISTRATION

Article

Survey experiments in Brazilian public administration: concepts, methods, and examples

João V. Guedes-Neto [1] *

* Corresponding Author

[1] Escola Brasileira de Administração Pública e de Empresas, Fundação Getulio Vargas, Rio de Janeiro, RJ, Brazil

Abstract

Survey experiments have become a central methodological tool in public administration research, enabling the identification of causal relationships in controlled settings. Despite their growing use in the international literature, their application in Brazil remains limited, particularly in studies involving civil servants. This article provides a practical guide to implementing survey experiments in Brazilian public administration, discussing their advantages, challenges, and methodological best practices. It presents three types of experiments widely used in the field: list experiments, which reduce social desirability bias when measuring sensitive attitudes and behaviors; vignette experiments, which examine how controlled variations in hypothetical scenarios influence perceptions and judgments; and fact-checking experiments, which assess belief updating following exposure to accurate information. For each experimental design, the article provides empirical examples applied in Brazil and discusses their underlying causal logic. It also outlines a step-by-step approach to formulating hypotheses, designing the questionnaire, collecting data, and statistically analyzing the results. The article contributes to the dissemination of experimental methods in the field and to strengthening evidence-based research in Brazilian public administration.
Keywords: survey experiments, experimental methods, public administration, civil servants, quantitative research.

Experimentos de survey na administração pública brasileira: conceitos, métodos e exemplos

Resumo

Os experimentos de *survey* têm se consolidado como uma ferramenta metodológica central na pesquisa em administração pública ao permitir a identificação de relações causais em contextos controlados. Apesar de seu uso crescente na literatura internacional, sua aplicação no Brasil ainda é limitada, especialmente em estudos que envolvem servidores públicos. Este artigo oferece um guia prático para a implementação de experimentos de *survey* na administração pública brasileira, discutindo suas vantagens, desafios e boas práticas metodológicas. São apresentados três tipos de experimentos amplamente utilizados no campo: experimentos de lista, que reduzem o

ISSN: 1982-3134

CC BY

DOI: https://doi.org/10.1590/0034-761220250695x
Submitted on December 31, 2025 and accepted on June 03, 2026.
[Translated version]
Editor-in-chief:
Gregory Michener, Fundação Getulio Vargas, Rio de Janeiro, RJ, Brazil
Associate editor:
Gustavo M. Tavares, Insper Instituto de Ensino e Pesquisa, São Paulo, SP, Brazil
Reviewers:
Marcel Ferreira de Oliveira, Universidade de São Paulo, São Paulo, SP, Brazil
Rodolfo de Camargo Lima, Universidad San Sebastián, Santiago, Chile
Peer review report:
The peer review report is available at https://periodicos.fgv.br/rap/article/view/97069/90457

viés de desejabilidade social na mensuração de atitudes e comportamentos sensíveis; experimentos de vinheta, que analisam como variações controladas em cenários hipotéticos influenciam percepções e julgamentos; e experimentos de checagem de fatos, que avaliam a atualização de crenças diante da exposição a informações corretas. Para cada desenho experimental, o artigo apresenta exemplos empíricos aplicados no Brasil e discute sua lógica causal. Além disso, detalha um passo a passo para a formulação de hipóteses, construção do questionário, coleta de dados e análise estatística dos resultados. O artigo contribui para a difusão do uso de métodos experimentais no campo e para o fortalecimento de pesquisas baseadas em evidências na administração pública brasileira.
Palavras-chave: experimentos de *survey*, métodos experimentais, administração pública, servidores públicos, pesquisa quantitativa.

Experimentos de encuesta en la administración pública brasileña: conceptos, métodos y ejemplos

Resumen

Los experimentos de encuesta se han consolidado como una herramienta metodológica central en la investigación sobre administración pública, ya que permiten identificar relaciones causales en contextos controlados. A pesar de su uso creciente en la literatura internacional, su aplicación en Brasil sigue siendo limitada, especialmente en estudios que involucran a servidores públicos. Este artículo ofrece una guía práctica para la implementación de experimentos de encuesta en la administración pública brasileña y analiza sus ventajas, desafíos y buenas prácticas metodológicas. Se presentan tres tipos de experimentos ampliamente utilizados en el campo: los experimentos de lista, que reducen el sesgo de deseabilidad social en la medición de actitudes y comportamientos sensibles; los experimentos de viñeta, que analizan cómo las variaciones controladas en escenarios hipotéticos influyen en las percepciones y los juicios; y los experimentos de verificación de hechos, que evalúan la actualización de creencias tras la exposición a información verídica. Para cada diseño experimental, el artículo presenta ejemplos empíricos aplicados en Brasil y analiza su lógica causal. Asimismo, detalla el paso a paso para la formulación de hipótesis, la elaboración del cuestionario, la recopilación de datos y el análisis estadístico de los resultados. El artículo contribuye a la difusión de los métodos experimentales en el campo y al fortalecimiento de la investigación basada en evidencias en la administración pública brasileña.
Palabras clave: experimentos de encuesta, métodos experimentales, administración pública, servidores públicos, investigación cuantitativa.

1. INTRODUCTION

Survey experiments allow researchers to causally test phenomena of interest in the public affairs field. Using this tool, it is possible to isolate and measure the effect of different interventions or individual- or institutional-level characteristics (Kertzer & Renshon, 2022; Mutz, 2011; Peters & Guedes-Neto, 2020), such as human resource policies (Sousa, 2024; Zimmermann, 2025), political appointments (Oliveros, 2021a; Oliveros & Schuster, 2017), political interference in the workplace (Ames & Guedes-Neto, 2025), democratic governance (Guedes-Neto & Peters, 2021, 2025), among others.

In this sense, the growing adoption of experimental methods in the field reflects the advancement of behavioral public administration (Battaglio et al., 2019; Bhanot & Linos, 2020; Grimmelikhuijsen et al., 2017; Kasdan, 2020), which emphasizes the importance of individual behavior for the functioning of the public sector. This movement, already encouraged by Simon (1997) in the first half of the last century, also accompanies a broader trend within the social sciences, in which experiments are used to overcome the limitations of observational research and establish more robust causal inferences (Sniderman, 2011).

However, despite the growing use of experimental methods in the international public administration literature (Bouwman & Grimmelikhuijsen, 2016; Hansen & Tummers, 2020; James et al., 2017; Jilke & Van Ryzin, 2017), their application in Brazil remains limited, particularly in

studies involving civil servants. One of the main reasons for this gap is the lack of familiarity among researchers and managers with the design, implementation, and analysis of survey experiments. In addition, the institutional context of the Brazilian public sector poses specific challenges, such as bureaucratic barriers to accessing samples of civil servants and potential resistance to participating in experimental research (Ames & Guedes-Neto, 2025).

This article aims to provide a practical guide to implementing survey experiments in Brazilian public administration, highlighting their advantages, challenges, and best practices. To this end, in addition to explaining the logic of this methodological tool, I present three of the main types of experiments frequently used in public administration research: list, vignette, and fact-checking experiments.

The choice of these three types does not imply that survey experiments are limited to them. Other designs can also be incorporated into questionnaires. The choice of list, vignette, and fact-checking experiments is based on three criteria: their relevance to recurring substantive questions in public administration; their pedagogical usefulness for introducing experimental logic; and their applicability to research with civil servants in contexts in which sensitive topics, judgments about institutional scenarios, and belief updating are particularly important.

In short, list experiments reduce social desirability bias when investigating opinions on sensitive topics, allowing respondents to express potentially controversial beliefs or behaviors without revealing their individual responses (Glynn, 2013; Kuklinski et al., 1997). Vignette experiments test how small variations in hypothetical scenarios influence respondents' perceptions and decisions (Alexander & Becker, 1978). Fact-checking experiments, in turn, assess the impact of correcting misinformation on respondents' beliefs and attitudes (Nyhan & Reifler, 2010), making them particularly relevant to studies of government transparency and organizational communication.

For each type of experiment, I present real examples applied in Brazil, showing how these methodologies can be used to investigate questions related to the functioning of the public sector. In addition, I provide a step-by-step guide to questionnaire construction and the statistical analysis of the results, ensuring that readers have a clear roadmap for applying these techniques.

I also emphasize that the use of survey experiments in public administration is not limited to academia. Public managers have increasingly shown interest in evidence-based approaches to improve policies and administrative processes. However, for this methodology to have a significant impact, it is essential to understand its limitations and challenges. Issues such as external validity and ethical considerations must be carefully assessed to ensure that the results obtained are robust and useful for public policy formulation (Peters & Guedes-Neto, 2020).

In this sense, at the end of this article, I discuss the implications of using survey experiments for Brazilian public administration and present recommendations for researchers and managers interested in adopting this approach. This article aims to contribute to the dissemination of this methodology in Brazil, promoting more rigorous research and helping to develop more effective, evidence-based public policies.

2. UNDERSTANDING SURVEY EXPERIMENTS

To understand the logic of survey experiments, one can use the classic analogy of testing a new drug. In a clinical trial, in simplified terms, two similar groups are created by randomizing a sample of patients. The first group receives the medication (treatment group), while the second receives a placebo (control group). At the end, the cure rates of patients in the two groups are compared. In short, if the treatment group has a higher cure rate than the control group, this difference can be attributed to the drug.

This comparison makes it possible to isolate the causal effect of the drug, that is, to identify whether it actually causes the observed improvement. This occurs because all other factors have been controlled through randomization of the sample, with the substance administered being the only difference between the groups.

This experimental logic can take different forms in the social sciences and applied social sciences. Laboratory experiments offer a high degree of control over the research environment, but they often involve more artificial situations and participants recruited outside their institutional contexts. Field experiments test interventions in real-world settings, more closely approximating the actual conditions of implementation, although they generally involve higher costs, less control over the context, and additional ethical and operational challenges. Natural experiments and quasi-experiments, in turn, exploit variations produced by events, rules, or policies external to the researcher, seeking to identify causal effects without the treatment having been directly manipulated by the research team.

Survey experiments occupy an intermediate position within this family: they preserve the controlled manipulation and randomization typical of experimental designs but embed the treatment in a questionnaire administered to a sample of respondents. In doing so, they allow researchers to test the causal effects of information, framing, or scenarios in a relatively simple, inexpensive, and scalable way, although their results must be interpreted in light of the limitations inherent in hypothetical responses, sampling, and external validity.

In this type of experiment, the logic is the same, but the "drug" is a piece of information, a scenario, or a way of presenting a public or organizational policy in a questionnaire. For example, one group may read a description of a policy stating that it was developed by technical staff, while another group reads that it was proposed by politicians. Because both groups answer the same questions after this intervention, the difference in their responses reveals the effect of this specific variation.

This comparison between what happened to the treated group and what would have happened to it had it not received the treatment is what we call counterfactual reasoning, that is, estimating an outcome that was not directly observed but inferred based on the control group. The major advantage of experimental methodology is that it introduces a counterfactual that would not be possible if our questionnaire were purely observational. This is possible because, by randomly assigning respondents to two groups, we can say that they are virtually identical. Thus, when one group is treated and the other is not, the control (untreated) group can be assumed to be the counterfactual of the treatment group.

In this sense, survey experiments combine two fundamental elements of scientific research: the principles of the classic experiment and the structure of surveys (or questionnaires administered to samples). From the classic experiment, they inherit the logic of controlled manipulation: researchers deliberately alter a variable (called the independent variable or treatment variable) and observe how this affects participants' responses (the dependent variable).

From surveys, this methodology takes advantage of the ability to administer questionnaires to large samples of people, allowing researchers to investigate causal effects in realistic settings and with different populations, such as civil servants or service users. With the advancement of survey platforms available to researchers (for example, Qualtrics, LimeSurvey, SurveyMonkey, and PsyToolkit), combined with basic quantitative training, it is possible to understand the growing number of studies in the international literature that use this tool.

This design makes it possible to estimate the so-called average treatment effect (ATE), which corresponds to the average difference in responses between the control group and the treatment group. This estimate can be obtained through statistical tests that, although relatively simple, are technically feasible and methodologically robust, such as the *t*-test, analysis of variance (ANOVA), or a linear regression with a dichotomous independent variable (Mutz, 2011; Shikano et al., 2012)—that is, using the ordinary least squares (OLS) method.

As Sniderman and Grob (1996) argue, this constitutes a "methodological revolution" that combined the internal validity advantages of experiments with the external validity of surveys, made possible primarily by the use of computer-assisted interviews and multifactorial manipulations through vignettes.

In public administration, survey experiments have proven particularly useful for investigating, among other topics, civil servants' preferences regarding meritocratic recruitment (Askim et al., 2024; Jankowski et al., 2020), perceptions of ethical behavior and motivation in the workplace (Ames & Guedes-Neto, 2025; Li & Yu, 2026; Meyer-Sahling et al., 2019), and the politics-administration relationship (Guedes-Neto & Peters, 2021, 2025; Lee & Park, 2020; Lobo et al., 2025). In these studies, rather than relying exclusively on the observation of correlational patterns in administrative data or generic self-reports in traditional surveys, researchers can isolate the causal effect of a variable by directly manipulating the content to which respondents are exposed and comparing similar groups created through the random assignment of respondents.

The main advantage of this method lies precisely in the control over the independent variable. In complex and multifactorial contexts such as public administration, establishing robust causal relationships requires separating the effect of the treatment (for example, the mention of a technical versus political attribute in a public policy) from other contextual or individual factors. Survey experiments allow this distinction by ensuring, through randomization, that any differences between treatment and control groups can be attributed exclusively to the manipulation performed (Sniderman, 2011; Mutz, 2011).

Another important benefit is the possibility of conducting causal inference in realistic settings. Unlike laboratory experiments, survey experiments can be administered on a large scale, using population samples or samples of specific groups (including civil servants) in their own institutional settings. This can increase the external validity of the results, especially when experimental scenarios are constructed in a plausible manner, as occurs in studies using vignettes (Alexander & Becker, 1978; Steiner et al., 2016).

In addition, the flexibility of the format makes it possible to adapt the design to different types of hypotheses. List experiments allow researchers to measure sensitive attitudes without directly exposing respondents (Ahlquist, 2018; Blair et al., 2020; Glynn, 2013); vignette experiments can manipulate attributes of a situation (such as the profile of a manager or the justifications for a policy) (Alexander & Becker, 1978; Gould, 1996; Taylor, 2006); and fact-checking experiments test the effectiveness of

misinformation corrections in changing beliefs (Nyhan & Reifler, 2010). These strategies provide precise tools for exploring not only whether an effect occurs, but also how and why it occurs, including through analyses of causal mechanisms.

However, despite these advantages, survey experiments are not a panacea and also have important limitations, particularly with regard to external validity and the interpretation of results in practical contexts (Peters & Guedes-Neto, 2020). As Barabas and Jerit (2010) observe, effects identified in survey experiments tend to be larger and clearer than those observed in natural settings, where individuals are exposed to multiple competing messages and contextual noise (Finkel & Geer, 1998).

For example, Guedes-Neto and Peters (2021, 2025) experimentally identify a high willingness among civil servants to sabotage the government when assigned to implement antidemocratic policies. Although this proposition has been confirmed in other studies (Bersch & Lotta, 2023; Lobo et al., 2025; Lotta et al., 2023), the real costs of resistance may mean that, in the day-to-day functioning of the bureaucracy, the proportion of civil servants willing to sabotage authoritarian policies is lower than that identified in the survey. For this reason, caution is needed when extrapolating average effects estimated in surveys to the real world.

Another challenge concerns the heterogeneity of treatment effects. As Coppock (2019) and Mullinix et al. (2015) show, even when nonprobability samples (such as convenience samples obtained online) produce effects similar to those found in representative samples, this occurs primarily when effects are homogeneous across subgroups. If there is a strong interaction between the treatment and unobserved characteristics of the population, the generalizability of the results may be compromised. In other words, it is valid to use convenience samples in survey experiments, but this should be done cautiously and under the assumption that the results will not necessarily be generalizable to the entire population. The literature provides some more sophisticated tests for estimating results in such cases, including the analysis of conditional average treatment effects (CATE).

Social desirability bias is another critical issue, especially when investigating politically sensitive or normatively charged topics, such as racial discrimination (Kuklinski et al., 1997) and vote buying (Gonzales-Ocantos et al., 2012). Respondents may adjust their answers to align with what they perceive as socially acceptable, distorting treatment effects or obscuring actual patterns of belief. The use of strategies such as list experiments or indirect manipulations helps address this problem but requires care in the construction and interpretation of the results (Blair et al., 2020; Glynn, 2013; Peters & Guedes-Neto, 2020).

From an ethical standpoint, survey experiments must respect the principles of informed consent, minimization of harm, and transparency regarding the nature of the study. In research with civil servants, these safeguards must be particularly explicit because participants may fear retaliation, institutional monitoring, or judgment of their responses. Therefore, the consent form should state, in plain language, that participation is voluntary, that refusal will have no professional consequences, that responses will be analyzed anonymously or in aggregate form, and that the research does not constitute an individual or institutional evaluation.

It is also advisable, when planning and reporting the results, to make the questionnaire, vignette texts, or experimental stimuli available whenever possible, and to register hypotheses and analysis strategies in advance when the research design and stage permit. As Harrits and Møller (2021) remind us, the sensitivity of the institutional context requires approaches that combine experimental control with interpretive sensitivity, while also valuing the use of qualitative vignettes in interviews to understand mechanisms of interpretation.

Finally, it is important to recognize that survey experiments do not replace other methods but complement them. Their main strength lies in identifying causal effects under controlled conditions. However, for them to have an impact on the formulation of public and organizational policies, it is essential to integrate their results with other empirical evidence, institutional knowledge, and an understanding of the practices and values of the actors involved. Thus, survey experiments should be viewed as part of a broader applied research strategy, with the potential to transform both academic scholarship and administrative practice in the Brazilian public sector.

3. FOUNDATIONS OF EXPERIMENTAL INFERENCE

3.1 Causal Inference and the Counterfactual Problem

The starting point of causal inference is the so-called counterfactual problem. To identify the causal effect of a policy, information, or intervention, it would be ideal to observe the same person in two absolutely identical parallel worlds, in which everything would be the same—preferences, institutional context, incentives, prior experiences—except for a single element: the presence or absence of the treatment. In one of these worlds, the individual would receive the treatment; in the other, they would not. The difference between the outcomes in these two scenarios would reveal the causal effect of the treatment.

This idea is frequently explored in science-fiction narratives based on parallel universes, such as the television series *Fringe*, created by J. J. Abrams, in which two worlds coexist with nearly identical histories but diverge as a result of small past interventions, generating different consequences over time. In the series, the characters encounter alternative versions of themselves and their trajectories, making it possible to observe how specific decisions produce different outcomes in contexts that would otherwise be equivalent.

This narrative exercise helps illustrate the central challenge of causal inference in the social sciences: in the real world, we do not have observable parallel universes. For each individual, organization, or government agency, only one outcome can be observed—the one that occurred under the condition actually experienced. The outcome that would have occurred under the alternative condition, that is, in the "other world" in which the treatment was not applied—or was applied differently—necessarily remains unobserved. This hypothetical outcome is what the literature calls the counterfactual, and the impossibility of observing it directly constitutes the fundamental problem of causal inference (Mutz, 2011).

In traditional observational surveys, this problem is addressed by comparing different individuals who differ with respect to the variable of interest. However, because individuals who receive different "treatments" in the real world also systematically differ across multiple unobserved dimensions, these comparisons frequently mix causal effects with selection bias (Mutz, 2011). Even with extensive statistical controls, uncertainty remains regarding the existence of unmeasured confounders, limiting the causal interpretation of the results. In simple terms, it is not possible to know whether the observed differences result from the treatment or from respondents' preexisting characteristics.

Survey experiments offer a crucial methodological advantage by addressing this problem directly. By randomizing treatment assignment, the researcher creates, in expectation, groups that are

equivalent across all dimensions—observed and unobserved—except for exposure to the treatment. In this way, the mean of the control group serves as a valid approximation of the counterfactual for the treated group (Mutz, 2011; Shikano et al., 2012). The comparison between groups ceases to be descriptive and acquires a direct causal interpretation, provided that randomization is successful and the experimental design is respected.

This counterfactual reasoning is particularly important in survey experiments applied to public administration. By manipulating scenarios, information, or framing presented to civil servants, the researcher does not observe actual behavior in complex institutional contexts but instead constructs plausible and controlled counterfactuals within the research instrument. As Acharya et al. (2018) emphasize, the value of survey experiments lies not in faithfully reproducing the real world but in isolating specific causal relationships under clearly defined conditions. This counterfactual understanding is precisely what distinguishes experiments from observational surveys and provides the basis for the causal interpretation of estimated effects.

3.2 Sampling and Randomization

Experimental research applied to public administration frequently faces substantial sampling constraints, especially when the focus is on civil servants and, in particular, street-level bureaucrats. These groups constitute hard-to-reach populations, whether because of institutional barriers, resistance to participation when sensitive topics are involved, or the limited size of the available empirical population.

In these contexts, the use of probability samples—that is, samples designed to statistically represent the entire population of interest—is rarely feasible. As a result, experiments in this area commonly rely on convenience samples composed of participants accessible to the researcher. As Krupnikov et al. (2021) argue, this type of sample does not allow researchers to describe precisely what the population as a whole is like, but this does not prevent its use for estimating causal effects in experiments, provided that the objective of the study and its limitations are made explicit.

The internal validity of experiments conducted with convenience samples depends primarily on the quality of randomization rather than on the degree to which the sample is representative. In simple terms, randomization—the random assignment of participants to control and treatment conditions—ensures that observed differences between groups can be attributed to the treatment rather than to participants' preexisting characteristics. It is therefore not necessary for the groups to be identical across all observable variables; it is sufficient for assignment to be random, producing groups that are comparable along theoretically relevant dimensions (Mutz, 2011).

In studies with civil servants, however, the randomization stage may be more challenging because participants frequently interact with one another in the workplace, exchange information, and may realize that they are taking part in a study. This can generate interference between groups, that is, situations in which the treatment applied to some participants indirectly affects others, requiring additional attention to the experimental design (Nathan & White, 2021). For researchers conducting their first studies, the main recommendation is simple: plan the randomization carefully and recognize that the organizational context matters for the interpretation of the results.

Finally, although convenience samples impose clear limits on the generalization of results—that is, on the possibility of claiming that the estimated effects apply to the entire civil service or to a

particular government agency—an extensive literature shows that experimental effects obtained with this type of sample are frequently replicated when the same experiment is conducted with probability samples. This pattern is especially common when the theoretical mechanisms being tested do not depend strongly on specific demographic characteristics of participants (Coppock, 2019; Coppock et al., 2018; Krupnikov et al., 2021).

In other words, even when a convenience sample is used and there is no guarantee of representativeness or generalizability, the estimated effects may be substantively similar to those observed in studies using more representative samples, provided that there is a plausible theoretical justification for expecting the mechanism being tested not to depend strongly on the specific characteristics of the sample.

More generally, in research with bureaucrats, the central point is not to eliminate all sampling limitations—which is often impossible—but to be transparent about the methodological choices made, acknowledge their limitations, and clearly discuss their analytical implications. This approach strengthens, rather than weakens, the credibility of experimental inference.

3.3 Experimental design

In survey experiments, experimental design refers to the way the researcher creates different versions of the questionnaire for different groups of respondents. The difference between these versions is called the treatment. In simple terms, the treatment is what changes from one condition to another—additional information, different framing, an alternative scenario, or a different way of presenting the same problem—while everything else in the survey remains the same. This deliberate variation is what makes it possible to compare groups and attribute differences in responses to the treatment received. Therefore, a change in the wording or version of a questionnaire constitutes a survey experiment only when it is associated with a clear definition of treatment and control, random assignment of respondents to conditions, and a transparent strategy for analyzing and reporting the results.

A central point emphasized by Mutz (2021) is that the treatment does not need to faithfully reproduce a real-world situation to be valid. The objective of the experiment is not to simulate perfectly how policies or decisions occur in practice, but to create a controlled difference between groups so that it is possible to observe how this difference affects participants' responses. In other words, the experiment seeks to answer the question: when we present different information or scenarios, do people respond differently? If the answer is yes, this difference can be interpreted causally, provided that the experimental design is respected.

In practice, a common error in survey experiments is the use of treatments that are excessively weak or insufficiently clear and therefore fail to generate perceptible differences between conditions. Even with proper randomization, treatments of this kind tend to produce small or nonexistent effects, making the results more difficult to interpret (Mutz, 2021). Researchers are therefore advised to be judicious in constructing treatments, seeking to make them understandable, relevant, and sufficiently distinct from one another.

In addition, the experimental design should be considered together with the structure of the questionnaire and the data-collection strategy. Whenever possible, the researcher should specify how respondents were recruited, what channel was used to access the sample, how many invitations were sent, how many responses were started and completed, which exclusion criteria were adopted,

and which limitations arise from nonresponse or from the inability to calculate a precise participation rate.

Before the main data collection, it is advisable to conduct a pretest or pilot of the instrument, even with a small number of participants, to assess comprehension of the treatments, the plausibility of the scenarios, the functioning of the randomization, and response time. In contexts such as research with bureaucrats—where formal pilots are not always feasible—a realistic strategy is to discuss the instrument in advance with colleagues in the field, subject-matter experts, or small groups of civil servants in order to assess whether the treatment makes sense and is interpreted as expected.

Finally, in institutional contexts where access to the sample is limited, sample planning should be compatible with the feasible number of respondents, avoiding designs that are excessively fragmented across many treatment groups and prioritizing focused hypotheses, clearly defined treatments, and analytical strategies proportionate to the sample size.

3.4 Questionnaire and data structure

In survey experiments, all participants begin the questionnaire by answering a common set of basic questions, such as gender, age, length of experience in the civil service, position, or area of work. This information serves two important purposes. First, it makes it possible to characterize respondents' professional and sociodemographic profiles. Second, it allows the sample obtained to be compared with the actual profile of the agency or bureaucracy under study, based on available administrative data—for example, through the Access to Information Law, the Personnel Statistical Panel, or institutional cooperation. This comparison does not make the sample representative, but it helps identify and discuss potential biases arising from convenience sampling, strengthening the analytical transparency of the study.

After this common block, participants are randomized between the control condition and the treatment condition. Historically, this required the use of separate questionnaires, which made proper randomization more difficult. Today, survey platforms make it possible to carry out this assignment simply and securely within a single instrument. Qualtrics, for example, offers a "survey flow" feature in which question blocks can be randomly assigned to respondents. Thus, after answering the same initial questions, some participants are exposed to the control content and others to the treatment content. Similar features are available on other platforms, such as SurveyMonkey and LimeSurvey (some requiring paid versions for advanced randomization), as well as free and open alternatives such as PsyToolkit.

It is important to emphasize that, in survey experiments, the difference between control and treatment lies not in the question itself, but in the text, scenario, or information presented immediately before it. The question, in turn, is identical and therefore repeated under both conditions. This makes it possible to compare the effect of different interventions on responses to the same question.

At the end of data collection, the data are organized in a spreadsheet using the standard survey format: each row represents a respondent (the observations), and each column represents a question (the variables). On platforms such as Qualtrics, it is common for the raw data to contain two columns associated with the experiment: one containing the dependent-variable responses for those in the control condition (with missing values for those in the treatment condition) and another containing the responses for those in the treatment condition (with missing values for those in the control

condition). Based on these columns, the researcher should construct two fundamental analytical variables: (i) a binary treatment-indicator variable, which takes the value 1 for respondents assigned to the treatment and 0 for those in the control group (the independent variable); and (ii) a single dependent variable that consolidates the observed responses across both conditions (the dependent variable). This structure is sufficient to apply, in any statistical software, the estimation procedures presented in the following section while keeping the counterfactual and experimental logic intact.

3.5 Estimating effects

In survey experiments with successful randomization, the *t*-test for differences in means, ANOVA, and linear regression with a categorical independent variable indicating treatment are fully acceptable statistical tools for estimating average causal effects. As Shikano et al. (2012) observe, the predominance of relatively simple tests in experimental studies does not result from methodological lag, but from the fact that randomization eliminates the main statistical problems that motivated the development of more sophisticated techniques for observational data, such as selection bias and systematic imbalance between groups.

In other words, in survey experiments, the main methodological investment takes place before the statistical analysis, in defining the treatment, constructing the instrument, and ensuring proper randomization. In this context, the *t*-test is appropriate for binary comparisons; ANOVA generalizes this reasoning to multiple groups; and linear regression represents an alternative parameterization of these same comparisons, producing statistically equivalent inferences when applied to the same experimental design.

Despite their inferential equivalence, these procedures differ in terms of analytical flexibility, statistical efficiency, and clarity in the presentation of results. The *t*-test for differences in means, as a bivariate procedure for comparing groups, is restricted to two groups and does not naturally extend to the inclusion of covariates or the analysis of treatment-effect heterogeneity. This should not be confused with the *t*-statistic used in regression models to test individual coefficients. ANOVA, in turn, although appropriate for designs with multiple treatments and interactions, offers less transparency in the substantive interpretation of estimated effects.

Linear regression—conceptually equivalent to ANOVA and analysis of covariance (ANCOVA)—allows covariates to be incorporated for the specific purpose of reducing residual variance and increasing statistical power, without this being necessary to ensure causal validity (Mutz, 2011; Shikano et al., 2012). As Mutz (2011) emphasizes, in well-designed experiments, the use of covariates should be parsimonious, theory-driven, and defined *ex ante*, since experimental logic already incorporates control over competing variables into the design itself. In this sense, regression is adopted in this study not as a correction for bias, but as a more general and flexible analytical tool.

In linear regression with an independent treatment-indicator variable and the control group defined as the reference category, the intercept (α) represents the expected mean of the control group, while the treatment coefficient (β) corresponds directly to the average treatment effect, that is, the difference between the mean of the treated group and the mean of the control group. The expected mean of the treated group is therefore given by $\alpha + \beta$, and the coefficient β can be substantively interpreted as the size of the causal effect of the treatment. In designs with multiple treatment groups, each coefficient indicates the difference between the mean of that specific group and the mean of the

reference group, while preserving the underlying experimental logic. This interpretation is consistent with the equivalence between regression, ANOVA, and ANCOVA highlighted by Shikano et al. (2012) and with Mutz's (2011) emphasis on analytical transparency and the clear presentation of means and differences in means as an appropriate way to communicate experimental results.

4. THREE TYPES OF SURVEY EXPERIMENTS

4.1 List experiments

List experiments, also known as item count techniques, are used to investigate sensitive attitudes or behaviors, especially when direct questions tend to generate underreporting due to social desirability bias. As described by Blair et al. (2020) and applied by Gonzalez-Ocantos et al. (2012), this method consists of presenting respondents with a list of statements and asking how many of them apply to them, without requiring them to specify which ones.

The key point is the variation between the control group and the treatment group. The first group receives a list of four items, while the second receives those same four items plus one additional item, with the additional item representing the sensitive context or behavior to be investigated.

The causal effect is estimated by the difference in the mean number of endorsed items between the control group (which sees a list without the sensitive item) and the treatment group (which sees the same list with the sensitive item added). This estimation is possible even if the researcher does not know which respondents actually selected this additional item in the treatment group. This makes it possible to measure attitudes that are difficult to capture through direct questions, particularly in organizational contexts marked by political or institutional risks.

Reducing social desirability bias is central to this experimental approach. Although survey experiments are generally administered anonymously, a considerable proportion of respondents may feel uncomfortable when asked to discuss sensitive topics. By indicating only the number of scenarios or behaviors in a list of items, without having to indicate which items they selected, respondents feel protected against any data leak or even against the negative feeling of directly revealing an undesirable preference.

Guedes-Neto and Peters (2021) conducted a list experiment with Brazilian municipal civil servants during the presidential transition from Michel Temer to Jair Bolsonaro to measure willingness to engage in bureaucratic resistance when faced with projects that would restrict democratic rights, especially through sabotage. In Brazil, the same experiment was replicated with public prosecutors at the National Treasury (PGFN) in Lobo's (2022) master's thesis—see also Lobo et al. (2025). Outside Brazil, it was replicated with American and British civil servants, respectively, after the election of Joe Biden—while Donald Trump was alleging electoral fraud—and during the period when Boris Johnson signed the United Kingdom's withdrawal from the European Union (Guedes-Neto & Peters, 2025).

In addition to confirming the willingness to resist, the list experiments identified that it is higher among civil servants without politically appointed positions and with greater perceived autonomy. These studies illustrate how the list method can capture forms of resistance in contexts of democratic erosion, overcoming the methodological challenges associated with social desirability bias.

The list experiment reported in this study follows this classic design. All respondents were exposed to the following preamble:

> The following four scenarios are common in public departments around the world. There is evidence that some of these scenarios reduce the motivation of several civil servants, causing them to devote less effort than they would to other activities. For example, they may try to pass the tasks on to other colleagues, perform them only partially, miss deadlines, or not do them. A civil servant was assigned to work on a project...

In the control group, respondents received a list of four scenarios common in public departments:

- ... similar to others on which the civil servant has always worked.
- ... exclusively favors their own political group.
- ... entirely new, requiring training and additional effort.
- ... creates a political advantage for groups that the civil servant opposes.

In the treatment group, these same four items were presented with the addition of a fifth sensitive scenario:

- ... reduces citizens' political freedoms, such as freedom of expression or freedom of the press.

The selection of these items follows well-established recommendations in the literature on list experiments, especially regarding the prevention of measurement errors associated with extreme responses. A known risk in this type of experiment occurs when respondents systematically choose the minimum or maximum possible value on the scale, generating floor effects or ceiling effects that can bias the estimate of the effect of the sensitive item (Ahlquist, 2018). To mitigate this problem, the literature suggests two complementary strategies.

The first consists of including, in the control list, at least one item that most respondents are expected to consider plausible or applicable, thereby avoiding excessive concentrations of responses at zero (Kuklinski et al., 1997). The second strategy consists of selecting items that are not strongly and positively correlated with one another—ideally, items whose selection makes the selection of other items less likely (Glynn, 2013). In the experiment presented here, these principles were incorporated by combining scenarios of opposing natures—political alignment, political misalignment, organizational cost, and administrative routine—thereby reducing the likelihood of extreme responses.

As a result of these choices, the mean number of endorsed items in the control group is expected to be close to the midpoint of the scale—approximately two in a list of four items. Although this condition is not strictly necessary for the validity of the experiment, it reduces the risk of measurement error and increases the precision of the estimate of the causal effect of the sensitive item (Ahlquist, 2018; Glynn, 2013). This design thus creates appropriate conditions for estimating the average treatment effect in the context of sensitive attitudes among civil servants.

Table 1 presents the estimation of the list experiment using a simple linear regression, in which the dependent variable (y) is the total number of endorsed items and the independent variable (x) is an indicator of the experimental condition (0 = control; 1 = treatment). As discussed in Section 3.5,

this specification is equivalent to a comparison of means: the intercept (α) corresponds to the mean of the control group, and the treatment coefficient (β) corresponds directly to the difference in means between treatment and control, that is, to the average effect of the sensitive item.

TABLE 1 LINEAR REGRESSION WITH A LIST EXPERIMENT ON BUREAUCRATIC RESISTANCE

	DV: Bureaucratic Resistance
Treatment	0.808***
	(0.177)
Intercept	1.959***
	(0.125)
Observations	146
R^2	0.127
Adjusted R^2	0.121

Note: *p < 0.1; **p < 0.05; ***p < 0.01.

Source: Guedes-Neto and Peters (2021, 2025).

In the control group, the estimated mean number of endorsed items was 1.96 (α = 1.9589; n = 73). This means that the objective of reaching the midpoint of the scale was achieved, thereby avoiding floor and ceiling effects. In the treatment group, the estimated mean was 2.77, obtained directly from α + β = 1.959 + 0.808 = 2.767 (n = 73). Thus, respondents exposed to the sensitive item endorsed, on average, 0.808 more items than respondents in the control group. This difference is statistically significant (β = 0.808; standard error [SE] = 0.177; p < 0.01), indicating that the inclusion of the sensitive item substantively increased the total number of scenarios considered capable of reducing the effort of a standard civil servant.

In a list experiment, this difference in means (treatment–control) can be interpreted, under the assumptions of the design, as an aggregate estimate of the proportion of respondents for whom the sensitive item "counts"—in other words, the proportion who, upon receiving the list with the additional item, would increase their count by 1 because they consider the sensitive scenario applicable. In the present case, the estimated effect was 0.808. This suggests that approximately 80.8% of respondents would endorse the sensitive item — "reduces citizens' political freedoms" — as a scenario capable of reducing the effort of a standard civil servant.

This interpretation, however, does not mean that we directly observed which individuals attributed this effect to the sensitive item. It is an aggregate estimate based on the comparison between the means of the treatment and control groups and dependent on the usual assumptions of the list experiment, especially the absence of design effects and the validity of aggregate responses. In practical terms, this is precisely why the method is useful: even without knowing who would individually select the sensitive item, comparing the group means makes it possible to estimate its endorsement in the sample as a whole.

In addition to the studies on bureaucratic resistance discussed in this section, list experiments have been used in research on sensitive behaviors that would be difficult to capture through direct questions. The classic study by Gonzalez-Ocantos et al. (2012) provides an example of the use of list experiments to measure vote buying in Nicaragua, demonstrating how direct questions can substantially underestimate socially sensitive behaviors. Oliveros (2016, 2021b), in turn, uses this strategy to investigate the relationship between politics and bureaucracy, considering issues such as clientelism and the influence of party machines. In a related area and with a focus on Brazil, Ames et al. (2025) use list experiments to investigate perceptions of political favoritism and fear of retaliation among state-level bureaucrats.

4.2 Vignette experiments

Vignette experiments are experimental designs used in surveys that present respondents with short, structured descriptions of hypothetical yet plausible situations involving individuals, organizations, or decisions. The central logic of this method is to hold the overall scenario constant while varying only one or a few specific elements across versions of the vignette randomly assigned to participants. As Alexander and Becker (1978) and Sniderman and Grob (1996) argue, this control over the stimulus presented reduces interpretive ambiguity and makes it possible to identify more precisely how specific variations affect respondents' judgments, evaluations, or decisions.

In practice, vignettes are particularly useful when researchers seek to investigate normative evaluations, subjective perceptions, or reactions to organizational dilemmas, in which direct responses to abstract questions tend to reflect heterogeneous mental images constructed by respondents themselves (Peters & Guedes-Neto, 2020). By explicitly describing the situation, the vignette aligns the "mental picture" to which individuals react, ensuring that observed differences in responses result from the experimental manipulation—and not from idiosyncratic interpretations of the problem or variables unobserved by the researcher (Alexander & Becker, 1978). This feature makes vignette experiments especially suitable for public administration research, in which decisions frequently involve complex trade-offs, conflicting norms, and specific institutional contexts.

Conceptually, vignette experiments may encompass what the literature calls framing effects. As Chong and Druckman (2007) define them, framing effects occur when small changes in the presentation of a problem produce substantive differences in individuals' opinions or evaluations. In vignettes, this framing is operationalized through the controlled alteration of attributes of the scenario—for example, the behavior of a supervisor, the justification presented for a decision, or the type of incentive employed—while holding everything else constant. In this way, the method makes it possible to test how different framings activate distinct criteria for judgment without altering the substantive context of the decision.

One of the main risks associated with vignette experiments is so-called satisficing, that is, the tendency of some respondents to process the stimulus superficially, without adequately reading or considering the scenario presented (Caro et al., 2012; Stolte, 1994). This behavior may generate treatment noncompliance or mistaken interpretations of the vignette, weakening the internal validity of the experiment (Harden et al., 2019). In such cases, the intention is to "treat" respondents, but some of them are not actually exposed to the treatment as planned.

To mitigate this problem, the literature recommends designs that encourage careful reading and engagement with the experimental stimulus, either through more engaging vignettes or through screen-time controls and manipulation-check questions that assess whether the respondent understood the central elements of the scenario (Harden et al., 2019). Although not every vignette experiment requires this level of methodological sophistication, explicitly acknowledging these risks and the strategies for addressing them helps strengthen the credibility of the results and analytical transparency.

In short, vignette experiments provide a powerful tool for isolating causal effects associated with specific attributes of organizational and decision-making scenarios. By combining experimental control, substantive plausibility, and operational flexibility, this method makes it possible to investigate how different framings, incentives, or justifications affect judgments in the public sector while holding the broader institutional context constant (Alexander & Becker, 1978; Chong & Druckman, 2007; Mutz, 2011; Peters & Guedes-Neto, 2020; Sniderman & Grob, 1996).

In Brazilian public administration, Cláudia Sousa (2024) applied a vignette experiment with civil servants from the National Social Security Institute (in Portuguese, Instituto Nacional do Seguro Social [INSS]) to test how different stimuli would affect perceptions of productivity and work quality in a context of delays in the analysis of applications. In the experiment, developed as part of her master's thesis, all participants were initially exposed to a common baseline text describing a plausible and recurring institutional situation in the INSS context:

> Imagine the following scenario: INSS decided to implement a new task force to address the problem of the queue of applications awaiting review. To this end, INSS decided to contact a group of civil servants responsible for reviewing these benefit applications. By email, INSS explained to the civil servants that, by increasing their productivity [...].

From this point onward, the additional content of the text—that is, the vignette itself—varied according to the experimental condition to which the respondent was randomly assigned. Each participant read only one of the three possible versions, which differed exclusively in the motivational framing associated with increased productivity, while everything else in the scenario remained constant. In the control group, increased productivity was associated exclusively with an organizational objective. In the first treatment group, the framing highlighted an intrinsic motivation associated with the institutional mission. In the second treatment group, the framing emphasized an extrinsic incentive of a financial nature.

- **Control**: "[...] they could reduce the queue of applications."
- **T1 (Social Function)**: "[...] they would help fulfill INSS's social role in serving citizens most in need."
- **T2 (Compensation)**: "[...] they would receive additional compensation."

After reading the vignette, all participants answered the same two questions, both measured on five-point Likert scales ranging from "Would decrease greatly" (1) to "Would increase greatly" (5). These two dependent variables constitute the central outcomes of the experiment. The first captured

expected productivity, while the second measured expected work quality. The questions were: In your opinion, what impact would this measure have on the number of cases analyzed by civil servants? In your opinion, what impact would this measure have on the quality of the analyses performed by civil servants?

From a methodological perspective, the design follows the classic logic of vignette experiments: all respondents evaluate exactly the same institutional scenario, differing only with respect to the experimentally manipulated attribute—the type of incentive associated with the additional effort. This strategy makes it possible to causally isolate how different motivational framings affect expectations regarding productivity and quality in the public service, while holding constant the organizational context, the content of the task, and the target population of the policy.

At the same time, the experiment has high empirical relevance. At the time of data collection, INSS faced a substantial backlog of applications, and one of the strategies actually adopted by the administration was the introduction of variable financial incentives to increase civil servants' productivity. This policy, however, was subject to internal criticism and receives mixed support in the literature on public sector motivation (Sousa, 2024). In this sense, the vignette experiment makes it possible to test, in a controlled manner and in direct connection with a real public management problem, how different justifications—organizational, extrinsic, or based on the institutional mission—shape expectations regarding performance.

Table 2 presents the results of the vignette experiment estimated using simple linear regressions, in which the dependent variables are, respectively, expected productivity (Quantitative) and expected quality (Qualitative). The independent variable is a categorical indicator of the experimental condition, with the control group as the reference category. As discussed in Section 3.5, this specification is equivalent to a direct comparison of means: the intercept (α) represents the expected mean in the control group, while the coefficients associated with the treatments (β) correspond to the mean differences between each treatment group and the control group.

TABLE 2 LINEAR REGRESSION WITH A VIGNETTE EXPERIMENT ON EXPECTATIONS OF PRODUCTIVITY AND QUALITY

	DV: Quantitative	DV: Qualitative
T1: Social Function	0.027	0.361***
	(0.099)	(0.090)
T2: Compensation	0.671***	-0.129
	(0.099)	(0.091)
Intercept	3.083***	2.284***
	(0.070)	(0.064)
Observations	657	657
R^2	0.083	0.046
Adjusted R^2	0.080	0.043

Note: *p < 0.1; **p < 0.05; ***p < 0.01.

Source: Sousa (2024).

In the case of expected productivity, the estimated mean in the control group was 3.08 (α = 3.083; $n \approx 218$), indicating that, in the absence of additional incentives, respondents expected, on average, a slight increase or maintenance of the productivity of the civil servants involved in the task force. The framing based on INSS's social function produced no statistically significant effect on this expectation (β = 0.027; SE = 0.099; $p = 0.788$), suggesting that simply evoking the institutional mission does not, on average, alter expectations regarding the number of cases analyzed.

In contrast, the framing based on additional compensation generated a substantively positive and statistically significant effect on expected productivity (β = 0.671; SE = 0.099; $p < 0.01$). This implies that respondents exposed to the financial incentive rated expected productivity, on average, approximately 0.67 points higher than the control group. The estimated mean for the compensation group was therefore approximately 3.75 (α + β = 3.083 + 0.671), indicating that extrinsic incentives are associated with higher expectations regarding productivity.

The results for the expected quality of the analyses follow a different pattern. In the control group, the estimated mean was 2.28 (α = 2.284; $n \approx 218$), indicating a relatively pessimistic assessment of the impact of the backlog of cases on work quality. Unlike the results for productivity, the framing based on INSS's social function had a positive and statistically significant effect on expected quality (β = 0.361; SE = 0.090; $p < 0.01$). This corresponds to an estimated mean of approximately 2.65 in the social-function group, suggesting that evoking the institutional mission consistently improves expectations regarding the quality of the analyses performed.

On the other hand, the framing based on additional compensation produced no statistically significant effect on expected quality (β = -0.129; SE = 0.091; $p = 0.154$). The negative sign of the coefficient even indicates a tendency—although statistically indistinguishable from zero—toward a reduction in perceived quality when additional effort is associated with financial incentives, compared with the control group.

Taken together, the results reveal a clear and theoretically consistent pattern: extrinsic financial incentives increase respondents' expectations regarding productivity, but not quality; framings based on the social impact of the work, in turn, increase expectations regarding quality without generating equivalent gains in expected productivity. This trade-off between quantity and quality is central to contemporary debates on performance management in the public sector and speaks directly to criticisms of the adoption of financial bonuses as an incentive instrument in public organizations (Sousa, 2024).

Other studies in the public affairs field also use vignette experiments to examine how small variations in institutional scenarios affect the judgments of civil servants or public managers. Migchelbrink and Van de Walle (2020), for example, investigate how participatory processes affect public officials' willingness to use societal input in administrative decisions. Pedersen et al. (2018) combine administrative data and a survey experiment to analyze how the ethnic origin of public service users influences discretionary sanctioning decisions by street-level bureaucrats in Danish employment agencies. In the Brazilian context, Guedes-Neto and Peters (2025) assess whether bureaucratic resistance is triggered merely by political disagreements or whether it depends on the perception of an attack on democracy. These studies illustrate the flexibility of the vignette design for investigating administrative judgments in situations in which formal rules, discretion, user characteristics, and forms of communication can alter respondents' interpretations.

4.3 Fact-checking experiments

Fact-checking experiments applied in surveys explicitly begin from the premise that individuals form judgments and attitudes based on prior beliefs (priors), which may be incomplete, inaccurate, or even incorrect (Nyhan & Reifler, 2010, 2015; Porter et al., 2018). In this type of experimental design, the control group receives no information and therefore answers the questions relying exclusively on these priors. The treatment group, by contrast, is exposed to verifiable and substantively correct information—whether about data, rules, policies, or organizations—making it possible to assess whether and how this information changes evaluations, perceptions, or attitudes.

The causal logic of the method is straightforward. In the absence of statistically significant effects, it is inferred that respondents already held well-informed beliefs about the object being evaluated or that the correction produced no difference in the measured outcome. Conversely, the presence of effects indicates that the information provided to the treatment group corrected previously inaccurate priors, systematically shifting participants' evaluations. Unlike vignette or list experiments, therefore, fact-checking experiments do not test latent preferences or normative judgments under different framings, but rather processes of belief updating in response to correct information.

In his master's thesis, Vinicius Zimmermann (2025) conducted a fact-checking experiment with civil servants from the Rio de Janeiro State Secretariat of Finance (in Portuguese, Secretaria de Estado de Fazenda do Rio de Janeiro [Sefaz-RJ]). The objective of the study was to assess whether civil servants actually knew the agency's institutional mission and, if not, how exposure to the formal mission statement would affect their organizational perceptions. The substantive premise of the experiment is that, in many organizational contexts, beliefs about institutional mission and performance are constructed informally through everyday work experience, internal interactions, and general perceptions of the organization—rather than through explicit knowledge of formal documents.

In the experiment, respondents were randomly assigned to two groups. The control group received no information before answering the questions and was therefore prompted to rely exclusively on its prior beliefs about the organization's mission, that is, by being asked to think about Sefaz-RJ's organizational mission without being explicitly told what that mission was. The treatment group, in contrast, was exposed to a paragraph containing the official mission statement, according to the institutional document in force at the time of the study:

> Guide, monitor, and evaluate the recording of acts and events related to the Budgetary, Financial, and Asset Management of State Public Administration bodies and entities, in order to produce information for decision-making by managers and for internal, external, and social oversight.

After reading the text (in the case of the treatment group), both groups evaluated two statements on numerical scales ranging from 0 to 10. The first measured individual subjective alignment with the institutional mission ("I feel aligned with Sefaz's mission"). The second captured perceptions of organizational capacity ("Sefaz is able to accomplish its mission"). These two variables constitute the central outcomes of the experiment.

From a methodological perspective, the comparison between the groups makes it possible to directly estimate the causal effect of the corrective information—that is, explicit knowledge of the organizational mission—on civil servants' evaluations. If respondents already had accurate knowledge of Sefaz's mission, exposure to the text should not produce any measurable effect.

The results indicate negative effects of exposure to the institutional mission on both outcomes analyzed. In the case of individual alignment with the mission, the intercept (α) estimates a mean of 7.52 points in the control group (SE = 0.26), while the treatment coefficient is negative and statistically significant (β = -0.77; SE = 0.38; p = 0.043), implying an estimated mean of approximately 6.75 points in the treatment group (α + β).

Similarly, for the perception that Sefaz is able to accomplish its mission, the control group has an estimated mean of 6.50 points (α = 6.50; SE = 0.26), whereas exposure to the mission statement reduces this evaluation by approximately 0.72 points (β = -0.72; SE = 0.38), resulting in a mean of approximately 5.78 points in the treatment group, with marginal statistical significance (p = 0.059).

These findings suggest that explicitly communicating the organizational mission did not reinforce civil servants' alignment or institutional confidence, but instead produced the opposite effect. The most plausible substantive interpretation, as discussed by Zimmermann (2025), is that civil servants held relatively positive—although imprecise—priors regarding Sefaz's mission. When confronted with the formal content of the mission, these priors were corrected, leading to more critical evaluations of both their own alignment and perceived organizational capacity.

From an analytical perspective, the experiment clearly illustrates the potential of fact-checking experiments to identify discrepancies between subjective beliefs and formal institutional information in the public sector. Rather than testing normative preferences or responses to incentives, this type of design makes it possible to examine how information affects organizational identities and perceptions of state capacity. The fact that Sefaz-RJ subsequently reformulated its institutional mission—although without any direct causal relationship with the study—reinforces the empirical relevance of the experiment and the importance of treating organizational communication as an object of causal investigation rather than as a taken-for-granted assumption.

Although still less common in public administration, fact-checking experiments have broad applications in studies of public opinion, political communication, and public policy. Nyhan and Reifler (2010) provide a classic example by testing whether corrections embedded in simulated news stories reduce political misperceptions, showing that corrections do not always eliminate false beliefs and may, in some cases, generate resistance. In a study applied to health policy, Nyhan et al. (2013) test whether corrections concerning U.S. health care reform reduce belief in the "death panels" myth. More recently, Aruguete et al. (2025) analyze how fact-checking messages in Argentina affect the perceived reputation of fact-checking organizations and their perceived ideological positioning. These examples show that this type of design is especially useful for investigating processes of belief updating in response to correct information, including on politically polarized or sensitive issues.

5. CONCLUSION

This article aimed to provide a practical guide to the use of survey experiments in public administration research, with a particular focus on the Brazilian context. Throughout the article, I argued that experiments constitute a methodological strategy particularly well suited to investigating central phenomena in the field—such as civil servants' perceptions, judgments, and behaviors—by allowing causal inferences with relatively low costs and a high degree of analytical control. Unlike exclusively observational approaches, this method makes it possible to isolate specific mechanisms, test substantive hypotheses transparently, and produce evidence directly relevant to the design and evaluation of public and organizational policies.

By presenting three types of experiments widely used in the literature—vignette, list, and fact-checking experiments—the article demonstrated how these tools can be operationalized in practice, from the formulation of treatments to the estimation and interpretation of effects. The empirical examples discussed, all applied in Brazil and involving actual civil servants, illustrate both the analytical potential of these designs and their substantive limitations.

List experiments made it possible to measure sensitive attitudes related to bureaucratic resistance; vignette experiments proved useful for identifying tensions between productivity and quality in the public service; and fact-checking experiments revealed processes of organizational belief updating in response to institutional information. Taken together, these cases reinforce that survey experiments not only work in the Brazilian context but also produce substantively informative and, at times, counterintuitive results.

Other experimental designs can also be incorporated into surveys and have gained prominence in the public administration literature and related fields. Among them are conjoint experiments, in which multiple attributes of profiles, policies, or scenarios are manipulated simultaneously, making it possible to estimate the relative importance of different components of a decision or judgment (Hainmueller et al., 2014). This design has been used, for example, to investigate how organizational attributes affect bureaucrats' willingness to resist antidemocratic pressures (Silveira et al., 2026) or how characteristics of an administrative order influence civil servants' propensity to engage in guerrilla government practices (Hollibaugh et al., 2019). Although these designs are not addressed in this guide, they represent a relevant extension for researchers interested in multidimensional choices and more complex administrative judgments.

In addition to its empirical contribution, the article sought to emphasize essential methodological best practices for conducting survey experiments in public administration. Issues such as treatment design, randomization, appropriate outcome selection, effect estimation, and substantive interpretation of results were addressed in an integrated manner, always with attention to the specific challenges of research involving civil servants. Ethical considerations, limitations of external validity, and potential institutional resistance were also discussed, reinforcing that the use of experiments does not eliminate the need for careful substantive judgment but instead requires even greater transparency and analytical rigor.

From the perspective of the field, the article argues that expanding the use of survey experiments can contribute to the advancement of Brazilian public administration, both in academic research and in public management practice. For researchers, this represents an opportunity to produce more robust causal inferences on topics traditionally studied descriptively. For public managers, survey experiments offer a relatively accessible tool for testing interventions, institutional messages, or organizational changes before their large-scale implementation, strengthening a culture of evidence-based policymaking.

As an introductory article, this text does not replace specialized manuals on experimental design and experimental methods in surveys. Researchers interested in further developing the planning, implementation, and analysis of these designs can consult more comprehensive works on experiments in the social sciences, such as Morton and Williams (2010) or Mutz (2011), as well as specific texts on survey, list, vignette, and fact-checking experiments cited throughout the article. The objective of this article is to provide an applied introduction tailored to Brazilian public administration, outlining

the main conceptual and operational considerations so that readers can subsequently move on to more specialized materials.

Finally, the article suggests that the future of survey experiments in public administration lies in their integration with other methodological approaches. Combining experiments with qualitative interviews, case studies, and network analyses can increase the external validity of the findings and deepen our understanding of the mechanisms underlying the observed effects. By promoting this plural methodological agenda, while anchoring it in rigorous causal inference, the article seeks to contribute to strengthening a research field that is more cumulative, empirically informed, and directly connected to the concrete challenges of the Brazilian public sector.

REFERENCES

Acharya, A., Blackwell, M., & Sen, M. (2018). Analyzing causal mechanisms in *survey* experiments. *Political Analysis*, *26*(4), 357-378. https://doi.org/10.1017/pan.2018.19

Ahlquist, J. S. (2018). List experiment design, non-strategic respondent error, and item count technique estimators. *Political Analysis*, *26*(1), 34-53. https://doi.org/10.1017/pan.2017.31

Alexander, C. S., & Becker, H. J. (1978). The use of vignettes in *survey* research. *Public Opinion Quarterly*, *42*(1), 93-104. https://doi.org/10.1086/268432

Ames, B., & Guedes-Neto, J. V. (Eds.). (2025). *Inside brazilian bureaucracy: politics and policy implementation in brazilian states*. Routledge.

Ames, B., Guedes-Neto, J. V., & McCoy, D. (2025). Bureaucrats and political bias. In B. Ames & J. V. Guedes-Neto (Eds.), *Inside brazilian bureaucracy: politics and policy implementation in brazilian states* (pp. 31-53). Routledge.

Aruguete, N., Calvo, E., & Ventura, T. (2025). The fact-checking dilemma: fact-checking increases the reputation of the fact-checker but creates perceptions of ideological bias. *Research & Politics*, *12*(1). https://doi.org/10.1177/20531680251323120

Askim, J., Bach, T., & Kolltveit, K. (2024). The professional profile, competence, and responsiveness of senior bureaucrats: a paired *survey* experiment with citizens and elite respondents. *Journal of Public Administration Research and Theory*, *35*(1), 87-101. https://doi.org/10.1093/jopart/muae024

Barabas, J., & Jerit, J. (2010). Are survey experiments externally valid? *American Political Science Review*, *104*(2), 226-242. https://doi.org/10.1017/S0003055410000092

Battaglio, R. P., Jr., Belardinelli, P., Bellé, N., & Cantarelli, P. (2019). Behavioral public administration *ad fontes*: a synthesis of research on bounded rationality, cognitive biases, and nudging in public organizations. *Public Administration Review*, *79*(3), 304-320. https://doi.org/10.1111/puar.12994

Bersch, K., & Lotta, G. (2023). Political control and bureaucratic resistance: the case of environmental agencies in Brazil. *Latin American Politics and Society*, *66*(1), 27–50. https://doi.org/10.1017/lap.2023.22

Bhanot, S. P., & Linos, E. (2020). Behavioral public administration: past, present, and future. *Public Administration Review*, *80*(1), 168-171. https://doi.org/10.1111/puar.13129

Blair, G., Coppock, A., & Moor, M. (2020). When to worry about sensitivity bias: a social reference theory and evidence from 30 years of list experiments. *American Political Science Review*, *114*(4), 1297-1315. https://doi.org/10.1017/S0003055420000374

Bouwman, R., & Grimmelikhuijsen, S. (2016). Experimental public administration from 1992 to 2014: a systematic literature review and ways forward. *International Journal of Public Sector Management*, *29*(2), 110-131. https://doi.org/10.1108/IJPSM-07-2015-0129

Caro, F. G., Ho, T., McFadden, D., Gottlieb, A. S., Yee, C., Chan, T., & Winter, J. (2012). Using the internet to administer more realistic vignette experiments. *Social Science Computer Review*, *30*(2), 184-201. https://doi.org/10.1177/0894439310391376

Chong, D., & Druckman, J. N. (2007). Framing public opinion in competitive democracies. *American Political Science Review*, 101(4), 637-655. https://doi.org/10.1017/S0003055407070554

Coppock, A. (2019). Generalizing from *survey* experiments conducted on mechanical turk: a replication approach. *Political Science Research and Methods*, *7*(3), 613-628. https://doi.org/10.1017/psrm.2018.10

Coppock, A., Leeper, T. J., & Mullinix, K. J. (2018). Generalizability of heterogeneous treatment effect estimates across samples. *Proceedings of the National Academy of Sciences*, *115*(49), 12441-12446. https://doi.org/10.1073/pnas.1808083115

Finkel, S. E., & Geer, J. G. (1998). A spot check: casting doubt on the demobilizing effect of attack advertising. *American Journal of Political Science*, *42*(2), 573-595. https://doi.org/10.2307/2991771

Glynn, A. N. (2013). What can we learn with statistical truth serum? Design and analysis of the list experiment. *Public Opinion Quarterly*, *77*(S1), 159-172. https://doi.org/10.1093/poq/nfs070

Gonzales-Ocantos, E., Jonge, C. K., Meléndez, C., Osorio, J., & Nickerson, D. W. (2012). Vote buying and social desirability bias: experimental evidence from Nicaragua. *American Journal of Political Science*, *56*(1), 202-217. https://doi.org/10.1111/j.1540-5907.2011.00540.x

Gould, D. (1996). Using vignettes to collect data for nursing research studies: how valid are the findings? *Journal of Clinical Nursing*, *5*(4), 207-212. https://doi.org/10.1111/j.1365-2702.1996.tb00253.x

Grimmelikhuijsen, S., Jilke, S., Olsen, A. L., & Tummers, L. (2017). Behavioral public administration: combining insights from public administration and psychology. *Public Administration Review*, *77*(1), 45-56. https://doi.org/10.1111/puar.12609

Guedes-Neto, J. V., & Peters, B. G. (2021). Working, shirking, and sabotage in times of democratic backsliding: an experimental study in Brazil. In M. W. Bauer, J. Pierre, K. Yesilkagit, B. G. Peters & S. Becker (Eds.), *Democratic backsliding and public administration: how populists in government transform state bureaucracies* (pp. 221-245). Cambridge University.

Guedes-Neto, J. V., & Peters, B. G. (2025). *Bureaucratic resistance in times of democratic backsliding*. Cambridge University.

Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (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

Hansen, J. A., & Tummers, L. (2020). A systematic review of field experiments in public administration. *Public Administration Review*, *80*(6), 921-931. https://doi.org/10.1111/puar.13181

Harden, J. J., Sokhey, A. E., & Runge, K. L. (2019). Accounting for noncompliance in survey experiments. *Journal of Experimental Political Science*, 6(3), 199-202. https://doi.org/10.1017/XPS.2019.13

Harrits, G. S., & Møller, M. Ø. (2021). Qualitative vignette experiments: a mixed methods design. *Journal of Mixed Methods Research*, *15*(4), 526-545. https://doi.org/10.1177/1558689820977607

Hollibaugh, G. E., Jr., Miles, M. R., & Newswander, C. B. (2019). Why public employees rebel: Guerrilla Government in the public sector. *Public Administration Review*, *80*(1), 64-74. https://doi.org/10.1111/puar.13118

James, O., Jilke, S. R., & Van Ryzin, G. G. (2017). Behavioural and experimental public administration: emerging contributions and new directions. *Public Administration*, *95*(4), 865-873. https://doi.org/10.1111/padm.12363

Jankowski, M., Prokop, C., & Tepe, M. (2020). Representative bureaucracy and public hiring preferences: evidence from a conjoint experiment among German municipal civil servants and private sector employees. *Journal of Public Administration Research and Theory*, *30*(4), 596-618. https://doi.org/10.1093/jopart/muaa012

Jilke, S., & Van Ryzin, G. G. (2017). Survey experiments for public management research. In O. James, S. R. Jilke, & G. G. Van Ryzin (Eds.), *Experiments in public management research: challenges and contributions* (pp. 117-138). Cambridge University. https://doi.org/10.1017/9781316676912.007

Kasdan, D. O. (2020). Toward a theory of behavioral public administration. *International Review of Administrative Sciences*, *86*(4), 605-621. https://doi.org/10.1177/0020852318801506

Kertzer, J. D., & Renshon, J. (2022). Experiments and surveys on political elites. *Annual Review of Political Science*, *25*, 529-550. https://doi.org/10.1146/annurev-polisci-051120-013649

Krupnikov, Y., Nam, H. H., & Style, H. (2021). Convenience samples in political science experiments. In J. N. Druckman & D. P. Green (Eds.), *Advances in Experimental Political Science* (pp. 165-183). Cambridge University.

Kuklinski, J. H., Cobb, M. D., & Gilens, M. (1997). Racial attitudes and the "New South". *Journal of Politics*, *59*(2), 323-349. https://doi.org/10.1017/S0022381600053470

Lee, D. S., & Park, S. (2020). Ministerial leadership and endorsement of bureaucrats: experimental evidence from presidential governments. *Public Administration Review*, *80*(3), 426-441. https://doi.org/10.1111/puar.13153

Li, Y., & Yu, H. (2026). Changes in the accountability obligation, intensity, and working drive of public employees: evidence from a survey experiment.

Journal of Public Administration Research and Theory, *36*(1), 71-84. https://doi.org/10.1093/jopart/muaf027

Lobo, J. R. F. (2022). *A resistência da burocracia brasileira diante de recuos democráticos: uma análise da PGFN* [Dissertação de mestrado, Fundação Getulio Vargas]. Repositório FGV. https://hdl.handle.net/10438/32233

Lobo, J. R. F., Marinho, M., Peci, A., & Guedes-Neto, J. V. (2025). Bureaucrats' willingness to resist democratic backsliding: an experimental study in the brazilian federal government. *Policy Studies*, 1-24. https://doi.org/10.1080/01442872.2025.2546382

Lotta, G. S., Lima, I. A., Fernandez, M., Silveira, M. C., Pedote, J., & Guaranha, O. L. C. (2023). A resposta da burocracia ao contexto de retrocesso democrático: uma análise da atuação de servidores federais durante o Governo Bolsonaro. *Revista Brasileira de Ciência Política*, (40), 1-36. https://doi.org/10.1590/0103-3352.2023.40.266094

Meyer-Sahling, J-H., Mikkelsen, K. S., & Schuster, C. (2019). The causal effect of public service motivation on ethical behavior in the public sector: evidence from a large-scale survey experiment. *Journal of Public Administration Research and Theory*, *29*(3), 445-459. https://doi.org/10.1093/jopart/muy071

Migchelbrink, K., & Van de Walle, S. (2020). When will public officials listen? A vignette experiment on the effects of input legitimacy on public officials' willingness to use public participation. *Public Administration Review*, *80*(2), 271-280. https://doi.org/10.1111/puar.13138

Morton, R. B., & Williams, K. C. (2010). *Experimental political science and the study of causality: from nature to the lab*. Cambridge University.

Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science*, *2*(2), 109-138. https://doi.org/10.1017/XPS.2015.19

Mutz, D. C. (2011). *Population-based survey experiments*. Princeton University.

Mutz, D. C. (2021). Improving experimental treatments in political science. In J. N. Druckman & D. P. Green (Eds.), *Advances in experimental political science* (pp. 219-238). Cambridge University.

Nathan, N. L., & White, A. (2021). Experiments on and with street-level bureaucrats. In J. N. Druckman & D. P. Green (Eds.), *Advances in experimental political science* (pp. 509-525). Cambridge University.

Nyhan, B., & Reifler, J. (2010). When corrections fail: the persistence of political misperceptions. *Political Behavior*, *32*, 303-330. https://doi.org/10.1007/s11109-010-9112-2

Nyhan, B., & Reifler, J. (2015). Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. *Vaccine*, *33*(3), 459-464. https://doi.org/10.1016/j.vaccine.2014.11.017

Nyhan, B., Reifler, J., & Ubel, P. A. (2013). The hazards of correcting myths about health care reform. *Medical Care*, *51*(2), 127-132. https://doi.org/10.1097/MLR.0b013e318279486b

Oliveros, V. (2016). Making it personal: clientelism, favors, and the personalization of public administration in Argentina. *Comparative Politics*, *48*(3), 373-391. https://doi.org/10.5129/001041516818254437

Oliveros, V. (2021a). *Patronage at work: public jobs and political services in Argentina*. Cambridge University.

Oliveros, V. (2021b). Working for the machine: patronage jobs and political services in Argentina. *Comparative Politics*, *53*(3), 381-427. https://doi.org/10.5129/001041521X15974977783469

Oliveros, V., & Schuster, C. (2017). Merit, tenure, and bureaucratic behavior: evidence from a conjoint experiment in the Dominican Republic. *Comparative Political Studies*, *51*(6), 759-792. https://doi.org/10.1177/0010414017710268

Pedersen, M. J., Stritch, J. M., & Thuesen, F. (2018). Punishment on the frontlines of public service delivery: client ethnicity and caseworker sanctioning decisions in a Scandinavian welfare state. *Journal of Public Administration Research and Theory*, *28*(3), 339-354. https://doi.org/10.1093/jopart/muy018

Peters, B. G., & Guedes-Neto, J. V. (2020). Experimental methods A: *survey* experiments in public administration. In E. Vigoda-Gadot & D. R. Vashdi (Eds.), *Handbook of research methods in public administration, management and policy* (pp. 218-233). Edward Elgar Publishing.

Porter, E., Wood, T. J., & Kirby, D. (2018). Sex Trafficking, Russian Infiltration, Birth Certificates, and Pedophilia: A Survey Experiment Correcting Fake News. *Journal of Experimental Political Science*, *5*(2), 159-164. https://doi.org/10.1017/XPS.2017.32

Shikano, S., Bräuninger, T., & Stoffel, M. (2012). Statistical analysis of experimental data. In B. Kittel, W. J. Luhan, & R. B. Morton (Eds.), *Experimental political science: principles and practices* (pp. 163-177). Palgrave Macmillan.

Silveira, M. C., Lotta, G. S., Cingolani, L., Guedes-Neto, J. V., Gomide, A. A., & Souza, P. M. S. (2026). The organizational dynamics of bureaucratic resistance to undemocratic pressures: a conjoint experiment in Brazil. *Public Administration Review*, 1-15. https://doi.org/10.1111/puar.70097

Simon, H. A. (1997). *Administrative Behavior* (4th ed.). Simon and Schuster.

Sniderman, P. M. (2011). The logic and design of the survey experiment: an autobiography of a methodological innovation. In J. N. Druckman, D. P. Green, J. H. Kuklinski, & A. Lupia (Eds.), *Cambridge Handbook of Experimental Political Science* (pp. 102-114). Cambridge University.

Sniderman, P. M., & Grob, D. B. (1996). Innovations in experimental design in attitude surveys. *Annual Review of Sociology*, *22*, 377-399. https://doi.org/10.1146/annurev.soc.22.1.377

Sousa, C. A. S. (2024). *A influência da remuneração variável na produtividade de analistas de processos de uma Autarquia Federal* [Dissertação de mestrado, Fundação Getulio Vargas, Escola Brasileira de Administração Pública e de Empresas]. Repositório FGV. https://hdl.handle.net/10438/36317

Steiner, P. M., Atzmüller, C., & Su, D. (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

Stolte, J. F. (1994). The context of satisficing in vignette research. *Journal of Social Psychology*, 134(6), 727-733. https://doi.org/10.1080/00224545.1994.9923007

Taylor, B. J. (2006). Factorial Surveys: using vignettes to study professional judgement. *The British Journal of Social Work*, *36*(7), 1187-1207. https://doi.org/10.1093/bjsw/bch345

Zimmermann, V. S. (2025). *Identificação organizacional: uma alternativa para gestão de pessoas no contexto do regime de recuperação fiscal no estado do Rio de Janeiro?* [Dissertação de Mestrado, Fundação Getulio Vargas, Escola Brasileira de Administração Pública e de Empresas]. Repositório FGV. https://hdl.handle.net/10438/36914

João V. Guedes-Neto

Ph.D. and M.A. in Political Science from the University of Pittsburgh; MA in Economics, Politics, and Law from Leuphana Universität Lüneburg; MA in Public Management and Society from the Universidade Federal de Alfenas; BA in Economics from the Universidade Federal de São João del-Rei; Assistant Professor of Public Administration at the Brazilian School of Public and Business Administration of the Fundação Getulio Vargas (FGV EBAPE); Co-editor of *Inside Brazilian Bureaucracy: Politics and Policy Implementation in Brazilian States* (Routledge); Co-author of *Bureaucratic Resistance in Times of Democratic Backsliding* (Cambridge University Press). E-mail: joao.neto@fgv.br

AUTHORS' CONTRIBUTION

João V. Guedes-Neto: Conceptualization (Lead); Data curation (Lead); Formal analysis (Lead); Investigation (Lead); Methodology (Lead); Project administration (Lead); Resources (Lead); Software (Lead); Supervision (Lead); Validation (Lead); Writing – original draft (Lead); Writing – review & editing (Lead).

RESEARCH DATA AVAILABILITY

The entire dataset supporting the results of this study is available upon request from the author. The dataset is not publicly available due to the inclusion of secondary data provided to the author for the purposes of this article.

CONFLICT OF INTEREST

The author declares no conflict of interest.

ARTIFICIAL INTELLIGENCE USAGE

The artificial intelligence tool ChatGPT was used to assist with the grammatical review of the text.