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

João V. Guedes-Neto

https://doi.org/10.1590/0034-761220250695

Submetido em: 2026-07-16
Postado em: 2026-07-17 (versão 1)
(AAAA-MM-DD)

RAP REVISTA DE ADMINISTRAÇÃO PÚBLICA

FGV EBAPE

Artigo

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

João V. Guedes-Neto

Fundação Getulio Vargas (FGV EBAPE), Escola Brasileira de Administração Pública e de Empresas, Rio de Janeiro, RJ, Brasil

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 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.

DOI: https://doi.org/10.1590/0034-761220250695
Submetido em 31 de dezembro de 2025 e aceito em 03 de junho de 2026.

Palavras-chave: experimentos de *survey*, métodos experimentais, administração pública, servidores públicos, pesquisa quantitativa.

Survey Experiments in Brazilian Public Administration: Concepts, Methods, and Examples

Abstract

Survey experiments have become a central methodological tool in public administration research because they enable 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 from Brazil and discusses its underlying causal logic. It also offers a step-by-step guide 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, public servants, quantitative research.

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 correcta. Para cada diseño experimental, el artículo presenta ejemplos empíricos aplicados en Brasil y analiza su lógica causal. Asimismo, ofrece una descripción detallada, paso a paso, de la formulación de hipótesis, la elaboración del cuestionario, la recopilación de datos y el análisis estadístico de los resultados. El artículo contribuye a difundir el uso de métodos experimentales en el campo y a fortalecer la investigación basada en evidencia en la administración pública brasileña.

Palabras clave: experimentos de encuesta, métodos experimentales, administración pública, servidores públicos, investigación cuantitativa.

1. INTRODUÇÃO

Experimentos de *survey* permitem testar de maneira causal fenômenos de interesse para pesquisadores do campo de públicas. A partir desta ferramenta, é possível isolar e mensurar o efeito de diferentes intervenções ou características de nível individual ou institucional (Kertzer & Renshon, 2022; Mutz, 2011; Peters & Guedes-Neto, 2020), como políticas de recursos humanos (Sousa, 2024; Zimmermann, 2025), nomeações políticas (Oliveros, 2021a; Oliveros & Schuster, 2017), interferência política no ambiente de trabalho (Ames & Guedes-Neto, 2025), governança democrática (Guedes-Neto & Peters, 2021, 2025), entre outras.

Nesse sentido, a adoção crescente de métodos experimentais no campo de pesquisa reflete o avanço da administração pública comportamental (Battaglio et al., 2019; Bhanot & Linos, 2020; Grimmelikhuijsen et al., 2017; Kasdan, 2020), que enfatiza a importância do comportamento individual para o funcionamento do setor público. Esse movimento, já incentivado por Simon (1997) na primeira metade do século passado, também acompanha uma tendência mais ampla dentro das ciências sociais, na qual os experimentos são utilizados

para superar limitações de pesquisas observacionais e estabelecer inferências causais mais robustas (Sniderman, 2011).

No entanto, apesar do crescimento do uso de métodos experimentais na literatura internacional de administração pública (Bouwman & Grimmelikhuijsen, 2016; Hansen & Tummers, 2020; James et al., 2017; Jilke & Van Ryzin, 2017), no Brasil sua aplicação ainda é limitada, especialmente em estudos que envolvem servidores públicos. Uma das principais razões para essa lacuna é a falta de familiaridade dos pesquisadores e gestores com o desenho, implementação e análise de experimentos de *survey*. Além disso, o contexto institucional do setor público brasileiro impõe desafios específicos, como barreiras burocráticas para acessar amostras de servidores e possíveis resistências à participação em pesquisas experimentais (Ames & Guedes-Neto, 2025).

Este artigo tem como objetivo fornecer um guia prático para a implementação de experimentos de *survey* na administração pública brasileira, destacando suas vantagens, desafios e boas práticas. Para isso, além de explicar a lógica desta ferramenta metodológica, apresento três dos principais tipos de experimentos frequentemente utilizados na pesquisa em administração pública: lista, vinheta e checagem de fatos.

A escolha desses três tipos não implica que os experimentos de *survey* se resumam a eles. Outros desenhos também podem ser incorporados a questionários. A opção por experimentos de lista, vinheta e checagem de fatos decorre de três critérios: sua relevância para questões substantivas recorrentes na administração pública; sua utilidade pedagógica para introduzir a lógica experimental; e sua aplicabilidade a pesquisas com servidores públicos em contextos nos quais temas sensíveis, julgamentos sobre cenários institucionais e atualização de crenças são particularmente importantes.

Em poucas palavras, experimentos de lista reduzem o viés de desejabilidade social ao investigar opiniões sobre temas sensíveis, permitindo que os respondentes expressem crenças ou comportamentos potencialmente controversos sem se exporem individualmente (Glynn, 2013; Kuklinski et al., 1997). Experimentos de vinheta testam como pequenas variações em cenários fictícios influenciam percepções e decisões dos respondentes (Alexander & Becker, 1978). Por sua vez, experimentos de checagem de fatos avaliam o impacto da correção de desinformação sobre crenças e atitudes dos respondentes (Nyhan & Reifler, 2010), sendo

particularmente relevantes para estudos sobre transparência governamental e comunicação organizacional.

Para cada tipo de experimento, apresento exemplos reais aplicados no Brasil, demonstrando como essas metodologias podem ser empregadas para investigar questões relacionadas ao funcionamento do setor público. Além disso, detalho um passo a passo para a construção do questionário e análise estatística dos resultados, garantindo que o leitor tenha um roteiro objetivo para a aplicação dessas técnicas.

Reforço, ainda, que o uso de experimentos de *survey* na administração pública não se limita ao campo acadêmico. Cada vez mais, gestores públicos têm se interessado por abordagens baseadas em evidências para aprimorar políticas e processos administrativos. No entanto, para que essa metodologia tenha impacto significativo, é fundamental compreender seus limites e desafios. Questões como validade externa e considerações éticas precisam ser cuidadosamente avaliadas para garantir que os resultados obtidos sejam robustos e úteis para a formulação de políticas públicas (Peters & Guedes-Neto, 2020).

Nesse sentido, discuto ao final deste artigo as implicações do uso de experimentos de *survey* para a administração pública brasileira e apresento recomendações para pesquisadores e gestores interessados em adotar essa abordagem. Este artigo pretende contribuir para a disseminação dessa metodologia no Brasil, promovendo pesquisas mais rigorosas e auxiliando na construção de políticas públicas mais eficazes e baseadas em evidências.

2. ENTENDENDO OS EXPERIMENTOS DE *survey*

Para entender a lógica dos experimentos de *survey*, é possível usar a analogia clássica dos testes de um novo remédio. Em um experimento clínico, de maneira simplificada, criam-se dois grupos parecidos a partir da aleatorização de uma amostra de pacientes. O primeiro grupo recebe o medicamento (grupo tratamento) e o segundo, um placebo (grupo controle). Ao fim, compara-se a taxa de cura dos pacientes dos dois grupos. Em síntese, se o grupo tratamento apresenta uma taxa de cura superior à do grupo controle, é possível atribuir essa diferença ao remédio.

Essa comparação permite isolar o efeito causal do remédio, ou seja, identificar se ele realmente causa a melhora observada. Isso acontece porque todos os demais fatores foram

controlados por meio da aleatorização da amostra, sendo que a única diferença entre os grupos é a substância administrada.

Essa lógica experimental pode assumir diferentes formatos no campo das ciências sociais e ciências sociais aplicadas. Experimentos de laboratório oferecem alto controle sobre o ambiente de pesquisa, mas frequentemente envolvem situações mais artificiais e participantes recrutados fora de seus contextos institucionais. Experimentos de campo testam intervenções em ambientes reais, aproximando-se mais das condições concretas de implementação, embora geralmente envolvam custos mais altos, menor controle sobre o contexto e desafios éticos e operacionais adicionais. Experimentos naturais e quase-experimentos, por sua vez, exploram variações produzidas por eventos, regras ou políticas externas ao pesquisador, buscando identificar efeitos causais sem que o tratamento tenha sido diretamente manipulado pela equipe de pesquisa.

Os experimentos de *survey* ocupam uma posição intermediária nessa família: preservam a manipulação controlada e a aleatorização típicas dos desenhos experimentais, mas inserem o tratamento em um questionário aplicado a uma amostra de respondentes. Com isso, permitem testar efeitos causais de informações, enquadramentos ou cenários de forma relativamente simples, barata e escalável, embora seus resultados devam ser interpretados à luz das limitações próprias de respostas hipotéticas, amostragem e validade externa.

Nesse tipo de experimento, a lógica é a mesma, mas o "remédio" é uma informação, um cenário ou uma forma de apresentar uma política pública ou organizacional em um questionário. Por exemplo, um grupo pode ler uma descrição de uma política dizendo que ela foi elaborada por técnicos, enquanto outro grupo lê que ela foi proposta por políticos. Sabendo que ambos os grupos responderão às mesmas perguntas após esta intervenção, a diferença nas respostas entre os dois grupos revela o efeito dessa variação específica.

Essa comparação entre o que aconteceu com o grupo tratado e o que teria acontecido com ele se não tivesse recebido o tratamento é o que chamamos de pensamento contrafactual, ou seja, estimar um resultado que não foi diretamente observado, mas inferido com base no grupo de controle. A grande vantagem da metodologia experimental é introduzir um contrafactual que não seria possível caso nosso questionário fosse puramente observacional. Isso acontece em virtude de, a partir da aleatorização dos respondentes entre dois grupos, podermos dizer que eles são virtualmente idênticos. Assim sendo, quando um grupo é tratado

e o outro não, pode-se assumir que o grupo controle (não tratado) é o contrafactual do grupo tratamento.

Nesse sentido, os experimentos de *survey* combinam dois elementos fundamentais da pesquisa científica: os princípios do experimento clássico e a estrutura dos *surveys* (ou questionários aplicados a amostras). Do experimento clássico, herda-se a lógica da manipulação controlada: os pesquisadores alteram deliberadamente uma variável (chamada de variável independente ou variável de tratamento) e observam como isso afeta as respostas dos participantes (a variável dependente).

Já dos *surveys*, essa metodologia aproveita a capacidade de aplicar questionários a grandes amostras de pessoas, permitindo investigar efeitos causais em ambientes realistas e com públicos variados, como servidores públicos ou usuários de serviços. Com o avanço das plataformas de *survey* disponíveis para pesquisadores (por exemplo, Qualtrics, LimeSurvey, SurveyMonkey e PsyToolkit), somado à formação básica quantitativa, é possível compreender o aumento do número de pesquisas na literatura internacional que se utilizam desta ferramenta.

Esse desenho permite estimar o chamado efeito médio do tratamento (do inglês *Average Treatment Effect* ou ATE), que corresponde à diferença média nas respostas entre o grupo controle e o grupo tratamento. Essa estimativa pode ser obtida por testes estatísticos que, embora relativamente simples, são tecnicamente viáveis e metodologicamente robustos como o teste t, a análise de variância (ANOVA) ou uma regressão linear com variável independente dicotômica (Mutz, 2011; Shikano et al., 2012) – ou seja, utilizando-se o método de mínimos quadrados ordinários (no inglês, *Ordinary Least Squares* ou OLS).

Como afirmam Sniderman e Grob (1996), trata-se de uma "revolução metodológica" que uniu as vantagens de validade interna dos experimentos à validade externa dos *surveys*, viabilizada sobretudo pelo uso de entrevistas computadorizadas e manipulações multifatoriais via vinhetas.

Na administração pública, os experimentos de *survey* têm se mostrado particularmente úteis para investigar, entre outros, preferências dos servidores públicos sobre recrutamento meritocrático (Askim et al., 2024; Jankowski et al., 2020), a percepção de comportamentos éticos e motivação no ambiente de trabalho (Ames & Guedes-Neto, 2025; Li & Yu, 2026; Meyer-Sahling et al., 2019) e a relação política-administração (Guedes-Neto

& Peters, 2021, 2025; Lee & Park, 2020; Lobo et al., 2025). Nessas pesquisas, ao invés de depender exclusivamente da observação de padrões correlacionais em dados administrativos ou de autorrelatos genéricos em *surveys* tradicionais, o pesquisador pode isolar o efeito causal de uma variável, manipulando diretamente o conteúdo a que o respondente é exposto e comparando grupos semelhantes criados a partir da aleatorização dos respondentes.

A principal vantagem desse método reside justamente no controle sobre a variável independente. Em contextos complexos e multifatoriais como a administração pública, estabelecer relações causais robustas exige separar o efeito do tratamento (por exemplo, a menção de um atributo técnico versus político em uma política pública) de outros fatores contextuais ou individuais. Os experimentos de *survey* permitem essa distinção ao garantir, via aleatorização, que quaisquer diferenças entre grupos de tratamento e controle possam ser atribuídas exclusivamente à manipulação realizada (Sniderman, 2011; Mutz, 2011).

Outro benefício importante é a possibilidade de realizar inferência causal em ambientes realistas. Diferentemente de experimentos de laboratório, os experimentos de *survey* podem ser aplicados em larga escala, por meio de amostras populacionais ou de grupos específicos (incluindo servidores públicos) em seus próprios ambientes institucionais. Isso pode aumentar a validade externa dos resultados, especialmente quando os cenários experimentais são construídos de forma verossímil, como ocorre nos estudos com vinhetas (Alexander & Becker, 1978; Steiner et al., 2016).

Além disso, a flexibilidade do formato permite adaptar o desenho a diferentes tipos de hipóteses. Experimentos de lista permitem medir atitudes sensíveis sem exposição direta do respondente (Ahlquist, 2018; Blair et al., 2020; Glynn, 2013); experimentos de vinheta podem manipular atributos de uma situação (como o perfil de um gestor ou as justificativas de uma política) (Alexander & Becker, 1978; Gould, 1996; Taylor, 2006); e experimentos de checagem de fatos testam a eficácia de correções de desinformação na mudança de crenças (Nyhan & Reifler, 2010). Essas estratégias oferecem instrumentos precisos para explorar não apenas se um efeito ocorre, mas como e por que ele ocorre, permitindo, inclusive, análises de mecanismos causais.

No entanto, apesar dessas vantagens, os experimentos de *survey* não são uma panaceia e também apresentam limitações importantes, especialmente no que diz respeito à validade externa e à interpretação dos resultados em contextos práticos (Peters & Guedes-

Neto, 2020). Como observam Barabas e Jerit (2010), efeitos identificados em *surveys* experimentais tendem a ser maiores e mais nítidos do que os observados em contextos naturais, nos quais os indivíduos são expostos a múltiplas mensagens concorrentes e ruídos contextuais (Finkel & Geer, 1998).

Por exemplo, Guedes-Neto e Peters (2021, 2025) identificam de maneira experimental uma alta disposição de servidores públicos a sabotar o governo quando designados a implementar políticas antidemocráticas. Ainda que esta proposição tenha sido confirmada em outros trabalhos (Bersch & Lotta, 2023; Lobo et al., 2025; Lotta et al., 2023), é possível que os custos reais da resistência façam com que, no dia a dia da burocracia, a proporção de servidores disposta a sabotar políticas autoritárias seja menor do que aquela identificada no *survey*. Por isso, é preciso cautela ao extrapolar os efeitos médios estimados em *surveys* para o mundo real.

Outro desafio diz respeito à heterogeneidade dos efeitos de tratamento. Como mostram Coppock (2019) e Mullinix et al. (2015), mesmo quando amostras não probabilísticas (como as de conveniência obtidas online) produzem efeitos semelhantes aos de amostras representativas, isso ocorre sobretudo quando os efeitos são homogêneos entre subgrupos. Se houver forte interação entre o tratamento e as características não observadas da população, a generalização dos resultados pode ser comprometida. Em outras palavras, é válido utilizar amostras de conveniência em experimentos de *survey*, mas isso deve ser feito com cautela e partindo da pressuposição de que os resultados não necessariamente serão generalizáveis para toda a população. A literatura oferece alguns testes mais sofisticados para estimar resultados nesses casos, incluindo a análise dos efeitos condicionais médios (no inglês, *Conditional Average Treatment Effects* ou CATE).

O viés de desejabilidade social é outro ponto crítico, especialmente quando se investigam temas politicamente sensíveis ou normativamente carregados, por exemplo, discriminação racial (Kuklinski et al., 1997) e compra de votos (Gonzales-Ocantos et al., 2012). Os respondentes podem ajustar suas respostas para se alinharem ao que percebem como socialmente aceitável, distorcendo os efeitos do tratamento ou obscurecendo padrões reais de crença. O uso de estratégias como experimentos de lista ou manipulações indiretas ajuda a contornar esse problema, mas exige cuidado na construção e na interpretação dos resultados (Blair et al., 2020; Glynn, 2013; Peters & Guedes-Neto, 2020).

Do ponto de vista ético, os experimentos de *survey* devem respeitar princípios de consentimento informado, minimização de danos e transparência sobre a natureza do estudo. Em pesquisas com servidores públicos, esses cuidados precisam ser particularmente explícitos, pois os participantes podem temer retaliações, monitoramento institucional ou julgamento de suas respostas. Por isso, o termo de consentimento deve informar, em linguagem simples, que a participação é voluntária, que a recusa não produzirá qualquer consequência profissional, que as respostas serão analisadas de forma anônima ou agregada e que a pesquisa não constitui avaliação individual ou institucional.

Também é recomendável, no planejamento e no reporte dos resultados, disponibilizar, quando possível, o questionário, os textos das vinhetas ou os estímulos experimentais; e registrar previamente hipóteses e estratégias de análise quando o desenho e o estágio da pesquisa permitirem. Como lembram Harrits & Møller (2021), a sensibilidade do contexto institucional exige abordagens que integrem controle experimental com sensibilidade interpretativa, valorizando também o uso de vinhetas qualitativas em entrevistas para compreender mecanismos de interpretação.

Por fim, é importante reconhecer que os experimentos de *survey* não substituem outros métodos, mas os complementam. Sua principal força reside na identificação de efeitos causais sob condições controladas. No entanto, para que tenham impacto na formulação de políticas públicas e organizacionais, é fundamental articular seus resultados com outras evidências empíricas, com conhecimento institucional e uma compreensão de práticas e valores dos atores envolvidos. Assim, os experimentos de *survey* devem ser vistos como parte de uma estratégia mais ampla de pesquisa aplicada, com potencial de transformar tanto a produção acadêmica quanto a prática administrativa no setor público brasileiro.

3. FUNDAMENTOS DA INFERÊNCIA EXPERIMENTAL

3.1 Inferência causal e o problema do contrafactual

O ponto de partida da inferência causal é o chamado problema do contrafactual. Para identificar o efeito causal de uma política, informação ou intervenção, seria ideal observar a mesma pessoa em dois mundos paralelos absolutamente idênticos, nos quais tudo fosse igual – preferências, contexto institucional, incentivos, experiências prévias –, exceto por um único

elemento: a presença ou ausência do tratamento. Em um desses mundos, o indivíduo receberia o tratamento; no outro, não. A diferença entre os resultados nesses dois cenários revelaria o efeito causal do tratamento.

Essa ideia é frequentemente explorada em narrativas de ficção científica baseadas em universos paralelos, como a série de televisão *Fronteiras* (no inglês, *Fringe*), criada por J. J. Abrams, na qual dois mundos coexistem com histórias quase idênticas, mas que divergem a partir de pequenas intervenções passadas, gerando consequências distintas ao longo do tempo. Na série, os personagens se deparam com versões alternativas de si mesmos e de suas trajetórias, permitindo observar como decisões pontuais produzem resultados diferentes em contextos que, de outra forma, seriam equivalentes.

Esse exercício narrativo ajuda a ilustrar o desafio central da inferência causal nas ciências sociais: no mundo real, não dispomos de universos paralelos observáveis. Para cada indivíduo, organização ou órgão público, só é possível observar um único resultado – aquele que ocorreu sob a condição efetivamente vivenciada. O resultado que teria ocorrido sob a condição alternativa, isto é, no "outro mundo" em que o tratamento não foi aplicado – ou foi aplicado de forma distinta —, permanece necessariamente não observado. Esse resultado hipotético é o que a literatura denomina contrafactual, e sua impossibilidade de observação direta constitui o problema fundamental da inferência causal (Mutz, 2011).

Em *surveys* observacionais tradicionais, esse problema é enfrentado por meio de comparações entre indivíduos distintos que diferem quanto à variável de interesse. No entanto, como indivíduos que recebem diferentes "tratamentos" no mundo real também diferem sistematicamente em múltiplas dimensões não observadas, essas comparações frequentemente misturam efeitos causais com vieses de seleção (Mutz, 2011). Mesmo com controles estatísticos extensivos, permanece a incerteza quanto à existência de confundidores não mensurados, limitando a interpretação causal dos resultados. Em termos simples, não é possível saber se as diferenças observadas decorrem do tratamento ou de características prévias dos respondentes.

Os experimentos de *survey* oferecem um ganho metodológico crucial ao lidar diretamente com esse problema. Ao randomizar a atribuição do tratamento, o pesquisador cria, em expectativa, grupos equivalentes em todas as dimensões – observáveis e não observáveis –, exceto pela exposição ao tratamento. Dessa forma, a média do grupo controle

funciona como uma aproximação válida do contrafactual do grupo tratado (Mutz, 2011; Shikano et al., 2012). A comparação entre grupos deixa de ser descritiva e passa a ter interpretação causal direta, desde que a randomização seja bem-sucedida e o desenho seja respeitado.

Esse raciocínio contrafactual é particularmente importante em experimentos de *survey* aplicados à administração pública. Ao manipular cenários, informações ou enquadramentos apresentados a servidores públicos, o pesquisador não observa comportamentos reais em contextos institucionais complexos, mas constrói contrafactuais plausíveis e controlados no instrumento de pesquisa. Como destacam Acharya et al. (2018), o valor dos experimentos de *survey* não está em reproduzir fielmente o mundo real, mas em isolar relações causais específicas sob condições nitidamente definidas. Essa compreensão contrafactual é justamente o que diferencia experimentos de *surveys* observacionais e fundamenta a interpretação causal dos efeitos estimados.

3.2 Amostragem e randomização

A pesquisa experimental aplicada à administração pública enfrenta, com frequência, restrições substantivas de amostragem, sobretudo quando o foco recai sobre servidores públicos e, em particular, sobre burocratas de nível de rua. Esses grupos constituem populações de difícil acesso (*hard-to-reach*), seja por barreiras institucionais, resistência à participação diante de temas sensíveis ou pelo tamanho limitado do universo empírico disponível.

Nesses contextos, o uso de amostras probabilísticas, isto é, amostras desenhadas para representar estatisticamente toda a população de interesse, é raramente viável. Como resultado, experimentos nessa área costumam recorrer a amostras de conveniência, formadas por participantes acessíveis ao pesquisador. Conforme argumentam Krupnikov et al. (2021), esse tipo de amostra não permite descrever com precisão como é a população como um todo, mas isso não impede seu uso para a estimação de efeitos causais em experimentos, desde que o objetivo do estudo e seus limites sejam explicitados.

A validade interna de experimentos conduzidos com amostras de conveniência depende, sobretudo, da qualidade da randomização, e não do grau de representatividade da

amostra. Em termos simples, a randomização – a alocação aleatória dos participantes entre as condições de controle e tratamento – garante que as diferenças observadas entre grupos possam ser atribuídas ao tratamento, e não a características prévias dos participantes. Não é necessário, portanto, que os grupos sejam idênticos em todas as variáveis observáveis; basta que a alocação seja feita de forma aleatória, produzindo grupos comparáveis nas dimensões teoricamente relevantes (Mutz, 2011).

Em estudos com servidores públicos, contudo, a etapa de randomização pode ser mais desafiadora, pois os participantes frequentemente interagem entre si no ambiente de trabalho, trocam informações e podem perceber que fazem parte de uma pesquisa. Isso pode gerar interferências entre grupos, isto é, situações em que o tratamento aplicado a alguns afeta indiretamente outros, exigindo atenção redobrada ao desenho do experimento (Nathan & White, 2021). Para pesquisadores que estão conduzindo seus primeiros estudos, a principal recomendação é simples: planejar cuidadosamente a randomização e reconhecer que o contexto organizacional importa para a interpretação dos resultados.

Por fim, embora amostras de conveniência imponham limites nítidos à generalização dos resultados, ou seja, à possibilidade de afirmar que os efeitos estimados valem para todo o serviço público ou para um determinado órgão do Estado, uma extensa literatura mostra que efeitos experimentais obtidos com esse tipo de amostra frequentemente se replicam quando o mesmo experimento é realizado com amostras probabilísticas. Esse padrão é especialmente comum quando os mecanismos teóricos testados não dependem fortemente de características demográficas específicas dos participantes (Coppock, 2019; Coppock et al., 2018; Krupnikov et al., 2021).

Em outras palavras, ainda que a amostra seja de conveniência e que não haja garantia de representatividade ou generalizabilidade, os efeitos estimados podem ser substantivamente semelhantes aos observados em estudos com amostras mais representativas, desde que haja uma justificativa teórica plausível para esperar que o mecanismo testado não dependa fortemente das características específicas da amostra.

De forma geral, no caso de pesquisas com burocratas, o ponto central não é eliminar todas as limitações de amostragem – o que muitas vezes é impossível –, mas ser transparente quanto às escolhas metodológicas realizadas, reconhecer seus limites e discutir

objetivamente suas implicações analíticas. Essa postura fortalece, e não enfraquece, a credibilidade da inferência experimental.

3.3 Desenho experimental

Em experimentos de *survey*, o desenho experimental refere-se à forma como o pesquisador cria versões diferentes do questionário para grupos distintos de respondentes. A diferença entre essas versões é chamada de tratamento. Em termos simples, o tratamento é aquilo que muda de uma condição para outra – uma informação adicional, um enquadramento distinto, um cenário alternativo ou uma forma diferente de apresentar o mesmo problema – enquanto todo o restante do *survey* permanece igual. Essa variação deliberada é o que permite comparar grupos e atribuir diferenças nas respostas ao tratamento recebido. Portanto, uma mudança de enunciado ou de versão do formulário só caracteriza um experimento de *survey* quando está associada a uma definição nítida de tratamento e controle, à alocação aleatória dos respondentes entre condições e a uma estratégia transparente de análise e reporte dos resultados.

Um ponto central enfatizado por Mutz (2021) é que o tratamento não precisa reproduzir fielmente uma situação do mundo real para ser válido. O objetivo do experimento não é simular com perfeição como políticas ou decisões ocorrem na prática, mas criar uma diferença controlada entre grupos, de modo que seja possível observar como esta afeta as respostas dos participantes. Em outras palavras, o experimento busca responder à pergunta: quando apresentamos informações ou cenários diferentes, as pessoas respondem de forma diferente? Se a resposta for sim, essa diferença pode ser interpretada causalmente, desde que o desenho experimental seja respeitado.

Na prática, um erro comum em experimentos de *survey* é o uso de tratamentos excessivamente fracos ou pouco objetivos, que não geram diferenças perceptíveis entre as condições. Mesmo com randomização correta, tratamentos desse tipo tendem a produzir efeitos pequenos ou inexistentes, dificultando a interpretação dos resultados (Mutz, 2021). Por isso, recomenda-se que pesquisadores sejam criteriosos na construção dos tratamentos, buscando torná-los compreensíveis, relevantes e suficientemente distintos entre si.

De maneira complementar, o desenho experimental deve ser pensado em conjunto com a estrutura do questionário e com a estratégia de coleta. Sempre que possível, o pesquisador deve explicitar como os respondentes foram recrutados, qual foi o canal de acesso à amostra, quantos convites foram enviados, quantas respostas foram iniciadas e concluídas, quais critérios de exclusão foram adotados e quais limitações decorrem de não resposta ou da impossibilidade de calcular uma taxa precisa de participação.

Antes da coleta principal, recomenda-se realizar um pré-teste ou piloto do instrumento, mesmo que com número reduzido de participantes, para verificar a compreensão dos tratamentos, a plausibilidade dos cenários, o funcionamento da randomização e o tempo de resposta. Em contextos como o estudo de burocratas – nos quais pilotos formais nem sempre são viáveis –, uma estratégia realista é discutir previamente o instrumento com colegas da área, especialistas no tema ou pequenos grupos de servidores, de modo a verificar se o tratamento faz sentido e é interpretado como esperado.

Por fim, em contextos institucionais nos quais o acesso à amostra é limitado, o planejamento amostral deve ser compatível com o número viável de respondentes, evitando desenhos excessivamente fragmentados em muitos grupos de tratamento e priorizando hipóteses focalizadas, tratamentos objetivos e estratégias de análise proporcionais ao tamanho da amostra.

3.4 Estrutura do questionário e dos dados

Em experimentos de *survey*, todos os participantes iniciam o questionário respondendo a um conjunto comum de perguntas básicas, como gênero, idade, tempo de experiência no serviço público, cargo ou área de atuação. Essas informações cumprem dois papéis importantes. Primeiro, permitem caracterizar o perfil profissional e sociodemográfico dos respondentes. Segundo, possibilitam comparar a amostra obtida com o perfil real do órgão ou da burocracia estudada, com base em dados administrativos disponíveis – por exemplo, via Lei de Acesso à Informação, Painel Estatístico de Pessoal ou por meio de cooperação institucional. Essa comparação não transforma a amostra em representativa, mas ajuda a identificar e discutir possíveis vieses decorrentes da amostragem por conveniência, fortalecendo a transparência analítica do estudo.

Após esse bloco comum, os participantes são aleatorizados entre a condição de controle e a condição de tratamento. Historicamente, isso exigia a aplicação de questionários distintos, o que dificultava a aleatorização adequada. Atualmente, plataformas de *survey* permitem realizar essa alocação de forma simples e segura dentro de um único instrumento. O Qualtrics, por exemplo, oferece a funcionalidade de "fluxo de pesquisa" (no inglês, *survey flow*), na qual blocos de perguntas podem ser atribuídos aleatoriamente aos respondentes. Com isso, após responderem às mesmas perguntas iniciais, parte dos participantes é exposta ao conteúdo de controle e outra parte ao conteúdo de tratamento. Funcionalidades semelhantes estão disponíveis em outras plataformas, como SurveyMonkey e LimeSurvey (algumas exigindo versões pagas para aleatorização avançada), além de alternativas gratuitas e abertas como o PsyToolkit.

É importante enfatizar que, nos experimentos de *survey*, a diferença entre controle e tratamento não está na pergunta em si, mas no texto, cenário ou na informação apresentada imediatamente antes dela. A pergunta, por sua vez, é idêntica e, portanto, repetida para as duas condições. Isso permite comparar o efeito de intervenções diferentes sobre a resposta de uma mesma pergunta.

Ao final da coleta, os dados são organizados em uma planilha no formato padrão de *surveys*: cada linha representa um respondente (as observações), e cada coluna representa uma pergunta (as variáveis). Em plataformas como o Qualtrics, é comum que o resultado bruto contenha duas colunas associadas ao experimento: uma com as respostas da variável dependente para quem esteve na condição de controle (com valores ausentes para o tratamento) e outra com as respostas para quem esteve no tratamento (com valores ausentes para o controle). Baseado nessas colunas, o pesquisador deve construir duas variáveis analíticas fundamentais: (i) uma variável binária indicadora de tratamento, que assume valor 1 para respondentes alocados no tratamento e 0 para o grupo de controle (a variável independente); e (ii) uma única variável dependente que consolida as respostas observadas em ambas as condições (a variável dependente). Essa estrutura é suficiente para aplicar, em qualquer *software* estatístico, os procedimentos de estimação apresentados na seção seguinte, mantendo a lógica contrafactual e experimental intacta.

3.5 Estimando efeitos

Em experimentos de *survey* com randomização bem-sucedida, o teste *t* de diferenças de médias, a ANOVA e a regressão linear com variável independente categórica indicadora de tratamento constituem instrumentos estatísticos plenamente aceitáveis para a estimação de efeitos causais médios. Como observam Shikano et al. (2012), a predominância de testes relativamente simples em estudos experimentais não decorre de atraso metodológico, mas do fato de que a randomização elimina os principais problemas estatísticos que motivaram o desenvolvimento de técnicas mais sofisticadas para dados observacionais, como viés de seleção e desequilíbrio sistemático entre grupos.

Em outras palavras, nos experimentos de *survey*, o principal investimento metodológico ocorre antes da análise estatística, na definição do tratamento, na construção do instrumento e na garantia de uma aleatorização adequada. Nesse contexto, o teste *t* é apropriado para comparações binárias; a ANOVA generaliza esse raciocínio para múltiplos grupos; e a regressão linear representa uma parametrização alternativa dessas mesmas comparações, produzindo inferências estatisticamente equivalentes quando aplicadas ao mesmo desenho experimental.

Apesar de sua equivalência inferencial, esses procedimentos diferem quanto à flexibilidade analítica, à eficiência estatística e à objetividade na apresentação dos resultados. O teste *t* de diferença de médias, como procedimento bivariado de comparação entre grupos, é restrito a dois grupos e não se estende naturalmente à inclusão de covariáveis ou à análise de heterogeneidade de efeitos. Isso não deve ser confundido com a estatística *t* utilizada em modelos de regressão para testar coeficientes individuais. Por sua vez, a ANOVA, embora adequada a desenhos com múltiplos tratamentos e interações, oferece menor transparência na interpretação substantiva dos efeitos estimados.

A regressão linear – conceitualmente equivalente à ANOVA e à Análise de Covariância (ANCOVA) – permite incorporar covariáveis com o objetivo específico de reduzir variância residual e aumentar poder estatístico, sem que isso seja necessário para garantir validade causal (Mutz, 2011; Shikano et al., 2012). Como enfatiza Mutz (2011), em experimentos bem desenhados, o uso de covariáveis deve ser parcimonioso, orientado por teoria e definido *ex ante*, uma vez que a lógica experimental já incorpora o controle de

variáveis concorrentes no próprio desenho. Nesse sentido, a regressão é adotada neste estudo não como correção de vieses, mas como ferramenta analítica mais geral e flexível.

Na regressão linear com variável independente indicadora de tratamento e categoria de referência definida como o grupo de controle, o intercepto (α) representa a média esperada do grupo controle, enquanto o coeficiente do tratamento (β) corresponde diretamente ao efeito médio do tratamento, isto é, à diferença entre a média do grupo tratado e a média do grupo controle. A média esperada do grupo tratado é, portanto, dada por $\alpha + \beta$, e o coeficiente β pode ser interpretado substantivamente como o tamanho do efeito causal do tratamento. Em desenhos com múltiplos grupos de tratamento, cada coeficiente indica a diferença entre a média daquele grupo específico e a média do grupo de referência, mantendo constante a lógica experimental subjacente. Essa interpretação é consistente com a equivalência entre regressão, ANOVA e ANCOVA destacada por Shikano et al. (2012) e com a ênfase de Mutz (2011) na transparência analítica e na apresentação objetiva de médias e diferenças de médias como forma adequada de comunicar resultados experimentais.

4. TRÊS TIPOS DE EXPERIMENTOS DE *SURVEY*

4.1 Experimentos de lista

Os experimentos de lista (*list experiments*), também conhecidos como técnicas de contagem de itens (*item count*), são empregados para investigar atitudes ou comportamentos sensíveis, especialmente quando perguntas diretas tendem a gerar subnotificação, em função do viés de desejabilidade social. Como descrito por Blair et al. (2020) e aplicado por Gonzalez-Ocantos et al. (2012), esse método consiste em apresentar aos respondentes uma lista de afirmações e perguntar quantas delas se aplicam a eles, sem exigir que especifiquem quais.

O ponto-chave é a variação que existe entre o grupo controle e o grupo tratamento. O primeiro grupo recebe uma lista de quatro itens, enquanto o segundo recebe esses mesmos quatro itens mais um, sendo que o item adicional é o contexto ou comportamento sensível que se pretende investigar.

O efeito causal é estimado pela diferença média no número de itens assinalados entre o grupo controle (que vê uma lista sem o item sensível) e o grupo tratamento (que vê a mesma lista acrescida do item sensível). Tal estimação se torna possível ainda que o pesquisador não saiba quais respondentes de fato escolheram este item adicional no grupo tratamento. Isso possibilita mensurar atitudes difíceis de captar por meio de perguntas diretas, sobretudo em contextos organizacionais marcados por riscos políticos ou institucionais.

Reduzir o viés de desejabilidade social é central para esta abordagem experimental. Ainda que experimentos de *survey* geralmente sejam aplicados de maneira anônima, uma parte considerável dos respondentes pode se sentir constrangida quando convidada a falar sobre temas sensíveis. Ao indicar apenas o número de cenários ou comportamentos em uma lista de itens sem precisar indicar os itens escolhidos, os respondentes se sentem protegidos contra qualquer vazamento de dados ou mesmo contra o sentimento negativo de revelar uma preferência indesejada de maneira direta.

Guedes-Neto e Peters (2021) conduziram um experimento de lista com servidores municipais brasileiros durante a transição presidencial de Michel Temer para Jair Bolsonaro a fim de mensurar a disposição à resistência burocrática diante de projetos que restringissem direitos democráticos, especialmente via sabotagem. No Brasil, o mesmo experimento foi replicado com Procuradores do Tesouro Nacional (PGFN) na dissertação de mestrado de Lobo (2022) – ver também Lobo et al. (2025). Já no exterior, houve replicação com servidores públicos americanos e britânicos, respectivamente, após a eleição de Joe Biden – enquanto Donald Trump acusava a existência de fraude eleitoral – e no período quando Boris Johnson assinava a saída do Reino Unido da União Europeia (Guedes-Neto & Peters, 2025).

Além de confirmarem a disposição à resistência, os experimentos de lista identificaram que ela é maior entre servidores sem cargos comissionados e com maior percepção de autonomia. Esses estudos ilustram como o método de lista permite captar formas de resistência em contextos de erosão democrática, superando os desafios metodológicos associados ao viés de desejabilidade social.

O experimento de lista reportado neste estudo segue esse desenho clássico. Todos os respondentes foram expostos ao seguinte preâmbulo:

> Os quatro cenários a seguir são comuns em repartições públicas ao redor do mundo. Há evidência que alguns destes cenários reduzem a motivação de diversos servidores públicos, fazendo com que eles dediquem menos esforços do que dedicariam para outras atividades. Por exemplo, eles podem tentar repassar as atividades para outros colegas, realizá-las apenas parcialmente, perder prazos, ou não as fazer. Um(a) servidor(a) público(a) foi designado(a) a trabalhar em um projeto...

No grupo controle, os respondentes receberam uma lista com quatro cenários comuns em repartições públicas:

- ... semelhante a outros nos quais o servidor sempre atuou.
- ... que favorece exclusivamente seu próprio grupo político.
- ... totalmente novo, que exige treinamento e esforço adicional.
- ... que cria vantagem política para grupos aos quais o servidor se opõe.

No grupo tratamento, esses mesmos quatro itens foram apresentados, acrescidos de um quinto cenário sensível:

- ... que reduz liberdades políticas dos cidadãos, como liberdade de expressão ou de imprensa.

A seleção desses itens segue recomendações consolidadas da literatura sobre experimentos de lista, especialmente no que diz respeito à prevenção de erros de mensuração associados a respostas extremas. Um risco conhecido desse tipo de experimento ocorre quando respondentes escolhem sistematicamente o valor mínimo ou o valor máximo possível da escala, gerando efeitos de piso (*floor effects*) ou de teto (*ceiling effects*) que podem enviesar a estimativa do efeito do item sensível (Ahlquist, 2018). Para mitigar esse problema, a literatura sugere duas estratégias complementares.

A primeira consiste em incluir, na lista de controle, ao menos um item que se espera que a maioria dos respondentes considere plausível ou aplicável, evitando concentrações

excessivas de respostas no valor zero (Kuklinski et al., 1997). A segunda estratégia consiste em selecionar itens que não sejam forte e positivamente correlacionados entre si – idealmente, itens cuja escolha torne menos provável a escolha de outros (Glynn, 2013). No experimento apresentado aqui, esses princípios foram incorporados por meio da combinação de cenários de natureza oposta – alinhamento político, desalinhamento político, custo organizacional e rotina administrativa –, reduzindo a probabilidade de respostas extremas.

Como resultado dessas escolhas, espera-se que, no grupo controle, a média do número de itens assinalados se situe próxima ao ponto médio da escala – aproximadamente dois, em uma lista de quatro itens. Embora essa condição não seja estritamente necessária para a validade do experimento, ela reduz o risco de erros de mensuração e aumenta a precisão da estimativa do efeito causal do item sensível (Ahlquist, 2018; Glynn, 2013). Esse desenho cria, assim, condições adequadas para a estimação do efeito médio do tratamento no contexto de atitudes sensíveis entre servidores públicos.

A Tabela apresenta a estimação do experimento de lista por meio de uma regressão linear simples, na qual a variável dependente (y) é o número total de itens assinalados e a variável independente (x) é um indicador da condição experimental (0 = controle; 1 = tratamento). Como discutido na Seção 3.5, essa especificação é equivalente a uma comparação de médias: o intercepto (α) corresponde à média do grupo controle, e o coeficiente do tratamento (β) corresponde diretamente à diferença de médias entre tratamento e controle, isto é, ao efeito médio do item sensível.

Tabela 1. REGRESSÃO LINEAR COM EXPERIMENTO DE LISTA SOBRE RESISTÊNCIA BUROCRÁTICA

	VD: Resistência Burocrática
Tratamento	0,808***
	(0,177)
Intercepto	1,959***
	(0,125)
Observações	146
R2	0,127
R2 Ajustado	0,121

Nota: *p < 0,1; **p < 0,05; ***p < 0,01.

Fonte: Guedes-Neto & Peters (2021; 2025).

No grupo controle, a média estimada do número de itens assinalados foi 1,96 (α = 1,9589; n = 73). Isso significa que o objetivo de atingir o ponto médio da escala foi atingido, portanto, evitando efeitos de piso e teto. No grupo tratamento, a média estimada foi 2,77, obtida diretamente por $\alpha + \beta$ = 1,959 + 0,808 = 2,767 (n = 73). Logo, os respondentes expostos ao item sensível assinalaram, em média, 0,808 item a mais do que os respondentes do grupo controle. Essa diferença é estatisticamente significativa (β = 0,808; Erro-Padrão [EP] = 0,177; p < 0,01), indicando que a inclusão do item sensível aumentou de forma substantiva o total de cenários considerados capazes de reduzir o empenho de um servidor público padrão.

Em um experimento de lista, essa diferença de médias (tratamento-controle) pode ser interpretada, sob as premissas do desenho, como uma estimativa agregada da proporção de respondentes para os quais o item sensível "conta" – em outras palavras, da parcela que, ao receber a lista com o item adicional, aumentaria sua contagem em 1 por considerar o cenário sensível aplicável. No presente caso, o efeito estimado foi 0,808. Isso sugere que aproximadamente 80,8% dos respondentes endossariam o item sensível – "reduz liberdades políticas dos cidadãos" – como um cenário capaz de reduzir o empenho de um servidor público padrão.

Essa interpretação, contudo, não significa que se tenha observado diretamente quais indivíduos atribuíram esse efeito ao item sensível. Trata-se de uma estimativa agregada, baseada na comparação entre médias dos grupos de tratamento e controle e dependente das premissas usuais do experimento de lista, especialmente a ausência de efeitos de desenho e a validade das respostas agregadas. Em termos práticos, é exatamente por isso que o método é útil: mesmo sem saber quem selecionaria o item sensível individualmente, a comparação entre as médias dos grupos permite estimar seu endosso no conjunto da amostra.

Além dos estudos sobre resistência burocrática discutidos nesta seção, experimentos de lista têm sido empregados em pesquisas sobre comportamentos sensíveis que dificilmente seriam captados por perguntas diretas. O estudo clássico de Gonzalez-Ocantos et al. (2012) oferece um exemplo do uso de experimentos de lista para mensurar compra de votos na

Nicarágua, demonstrando como perguntas diretas podem subestimar fortemente comportamentos socialmente sensíveis. Oliveros (2016, 2021b), por sua vez, utiliza essa estratégia para investigar a relação entre política e burocracia, considerando questões como clientelismo e o peso das máquinas partidárias. Em área correlata e com foco no Brasil, Ames et al. (2025) utilizam experimentos de lista para investigar percepções de favoritismo político e medo de represália entre burocratas estaduais.

4.2 Experimentos de vinheta

Os experimentos de vinheta (*vignette experiments*) são desenhos experimentais aplicados em *surveys* que apresentam aos respondentes descrições curtas e estruturadas de situações hipotéticas, porém plausíveis, envolvendo indivíduos, organizações ou decisões. A lógica central desse método é manter o cenário geral constante, variando apenas um ou poucos elementos específicos entre versões da vinheta atribuídas aleatoriamente aos participantes. Como argumentam Alexander e Becker (1978) e Sniderman e Grob (1996), esse controle sobre o estímulo apresentado reduz ambiguidades interpretativas e permite identificar de forma mais precisa como variações pontuais afetam julgamentos, avaliações ou decisões dos respondentes.

Na prática, vinhetas são particularmente úteis quando o pesquisador deseja investigar avaliações normativas, percepções subjetivas ou reações a dilemas organizacionais, nos quais respostas diretas a perguntas abstratas tendem a refletir imagens mentais heterogêneas construídas pelos próprios respondentes (Peters & Guedes-Neto, 2020). Ao descrever explicitamente a situação, a vinheta alinha o "quadro mental" ao qual os indivíduos reagem, garantindo que diferenças observadas nas respostas decorram da manipulação experimental – e não de interpretações idiossincráticas do problema ou de variáveis não observadas pelo pesquisador (Alexander & Becker, 1978). Esse aspecto torna os experimentos de vinheta especialmente adequados para pesquisas em administração pública, nas quais decisões frequentemente envolvem *trade-offs* complexos, normas conflitantes e contextos institucionais específicos.

Do ponto de vista conceitual, experimentos de vinheta podem englobar o que a literatura denomina efeitos de enquadramento (*framing effects*). Como definem Chong e

Druckman (2007), efeitos de enquadramento ocorrem quando pequenas mudanças na apresentação de um problema produzem diferenças substantivas nas opiniões ou avaliações dos indivíduos. Em vinhetas, esse enquadramento é operacionalizado por meio da alteração controlada de atributos do cenário – por exemplo, o comportamento de um superior hierárquico, a justificativa apresentada para uma decisão ou o tipo de incentivo mobilizado –, mantendo-se todo o restante constante. Dessa forma, o método permite testar como diferentes enquadramentos ativam critérios distintos de julgamento sem alterar o contexto substantivo da decisão.

Um dos principais riscos associados a experimentos de vinheta é o chamado *satisficing*, isto é, a tendência de alguns respondentes a processar o estímulo de forma superficial, sem ler ou considerar adequadamente o cenário apresentado (Caro et al., 2012; Stolte, 1994). Esse comportamento pode gerar não conformidade com o tratamento (*noncompliance*) ou interpretações equivocadas da vinheta, enfraquecendo a validade interna do experimento (Harden et al., 2019). Nesses casos, há a intenção de "tratar" os respondentes, mas parte deles não é efetivamente exposta ao tratamento conforme planejado.

Para mitigar esse problema, a literatura recomenda desenhos que incentivem a leitura atenta e o engajamento com o estímulo experimental, seja por meio de vinhetas mais envolventes, seja por meio de controles de tempo de tela e perguntas de checagem de manipulação que avaliem se o respondente compreendeu os elementos centrais do cenário (Harden et al., 2019). Ainda que nem todo experimento de vinheta exija esse nível de sofisticação metodológica, o reconhecimento explícito desses riscos e das estratégias para enfrentá-los contribui para fortalecer a credibilidade dos resultados e a transparência analítica.

Em síntese, experimentos de vinheta oferecem uma ferramenta poderosa para isolar efeitos causais associados a atributos específicos de cenários organizacionais e decisórios. Ao combinar controle experimental, plausibilidade substantiva e flexibilidade operacional, esse método permite investigar como diferentes enquadramentos, incentivos ou justificativas afetam julgamentos no setor público, mantendo constante o contexto institucional mais amplo (Alexander & Becker, 1978; Chong & Druckman, 2007; Mutz, 2011; Peters & Guedes-Neto, 2020; Sniderman & Grob, 1996).

Na administração pública brasileira, Cláudia Sousa (2024) aplicou um experimento de vinheta com servidores do Instituto Nacional do Seguro Social (INSS) para testar como diferentes estímulos afetariam a percepção de produtividade e qualidade do trabalho em um contexto de atrasos na análise de requerimentos. No experimento, desenvolvido no âmbito de sua dissertação de mestrado, todos os participantes foram inicialmente expostos a um texto-base comum, que descrevia uma situação institucional plausível e recorrente no contexto do INSS:

> Imagine o seguinte cenário, o INSS decidiu implementar uma nova força-tarefa para resolver o problema da fila de requerimentos que aguardam análise. Para isto, o INSS decidiu contactar um grupo de servidores responsável pela análise destes benefícios. Por e-mail, o INSS explicou aos servidores que, ao aumentar sua produtividade [...].

A partir desse ponto, o conteúdo complementar do texto, isto é, a vinheta propriamente dita, variava de acordo com a condição experimental à qual o respondente foi aleatoriamente alocado. Cada participante leu apenas uma das três versões possíveis, que diferiam exclusivamente quanto ao enquadramento motivacional associado ao aumento da produtividade, mantendo todo o restante do cenário constante. No grupo controle, o aumento da produtividade era associado exclusivamente a um objetivo organizacional. No primeiro grupo de tratamento, o enquadramento destacava uma motivação intrínseca associada à missão institucional. Já no segundo grupo de tratamento, o enquadramento enfatizava um incentivo extrínseco de natureza financeira.

- Controle: “[...] poderiam reduzir a fila de requerimentos”.
- T1 (Função Social): “[...] ajudariam no papel social do INSS em atender os cidadãos mais carentes”.
- T2 (Remuneração): “[...] receberiam uma remuneração adicional”.

Após a leitura da vinheta, todos os participantes responderam às mesmas duas perguntas, ambas medidas em escalas Likert de cinco pontos, variando de “Diminuiria

muito" (1) a "Aumentaria muito" (5). Estas duas variáveis dependentes constituem os desfechos (*outcomes*) centrais do experimento. A primeira captava a produtividade esperada, enquanto a segunda mensurava a qualidade esperada do trabalho. São elas: Na sua opinião, qual o impacto que esta medida teria na quantidade de processos analisados pelos servidores? Na sua opinião, qual o impacto que esta medida teria na qualidade das análises realizadas pelos servidores?

Do ponto de vista metodológico, o desenho segue a lógica clássica dos experimentos de vinheta: todos os respondentes avaliam exatamente o mesmo cenário institucional, diferindo apenas quanto ao atributo experimentalmente manipulado – o tipo de incentivo associado ao esforço adicional. Essa estratégia permite isolar de forma causal como diferentes enquadramentos motivacionais afetam expectativas sobre produtividade e qualidade no serviço público, mantendo constante o contexto organizacional, o conteúdo da tarefa e o público-alvo da política.

Ao mesmo tempo, o experimento possui elevada relevância empírica. À época da coleta de dados, o INSS enfrentava uma fila expressiva de requerimentos represados, e uma das estratégias efetivamente adotadas pela administração foi a introdução de incentivos financeiros variáveis para elevar a produtividade dos servidores. Essa política, contudo, foi alvo de críticas internas e encontra respaldo ambíguo na literatura sobre motivação no setor público (Sousa, 2024). Nesse sentido, o experimento de vinheta permite testar, de forma controlada e diretamente vinculada a um problema real de gestão pública, como diferentes justificativas – organizacionais, extrínsecas ou baseadas na missão institucional – moldam expectativas sobre desempenho.

A Tabela 1 apresenta os resultados do experimento de vinheta estimados por meio de regressões lineares simples, nas quais as variáveis dependentes são, respectivamente, a produtividade esperada (Quantitativo) e a qualidade esperada (Qualitativo). A variável independente é um indicador categórico da condição experimental, tendo o grupo controle como categoria de referência. Conforme discutido na Seção 3.5, essa especificação é equivalente a uma comparação direta de médias: o intercepto (α) representa a média esperada no grupo controle, enquanto os coeficientes associados aos tratamentos (β) correspondem às diferenças médias entre cada grupo de tratamento e o grupo controle.

Tabela 1. REGRESSÃO LINEAR COM EXPERIMENTO DE VINHETA SOBRE EXPECTATIVA DE PRODUTIVIDADE E QUALIDADE

	VD: Quantitativo	**VD: Qualitativo**
T1: Função Social	0,027	0,361***
	(0,099)	(0,090)
T2: Remuneração	0,671***	-0,129
	(0,099)	(0,091)
Intercepto	3,083***	2,284***
	(0,070)	(0,064)
Observações	657	657
R2	0,083	0,046
R2 Ajustado	0,080	0,043

Nota: *p < 0,1; **p < 0,05; ***p < 0,01.

Fonte: Sousa (2024).

No caso da produtividade esperada, a média estimada no grupo controle foi de 3,08 (α = 3,083; n ≈ 218), indicando que, na ausência de incentivos adicionais, os respondentes esperavam, em média, um leve aumento ou manutenção da produtividade dos servidores envolvidos na força-tarefa. O enquadramento baseado na função social do INSS não produziu qualquer efeito estatisticamente significativo sobre essa expectativa (β = 0,027; EP = 0,099; p = 0,788), sugerindo que a simples evocação da missão institucional não altera, em média, a expectativa sobre a quantidade de processos analisados.

Em contraste, o enquadramento baseado em remuneração adicional gerou um efeito positivo substantivo e estatisticamente significativo sobre a produtividade esperada (β = 0,671; EP = 0,099; p < 0,01). Isso implica que os respondentes expostos ao incentivo financeiro avaliaram, em média, a produtividade esperada em aproximadamente 0,67 ponto acima do grupo controle. A média estimada para o grupo de remuneração foi, portanto, de aproximadamente 3,75 ($\alpha + \beta$ = 3,083 + 0,671), evidenciando que incentivos extrínsecos estão associados a expectativas mais elevadas de produtividade esperada.

Os resultados para a qualidade esperada das análises seguem um padrão distinto. No grupo controle, a média estimada foi de 2,28 (α = 2,284; n ≈ 218), indicando uma avaliação

relativamente pessimista quanto ao impacto da fila de processos sobre a qualidade do trabalho. Diferentemente do observado para produtividade, o enquadramento baseado na função social do INSS teve um efeito positivo e estatisticamente significativo sobre a qualidade esperada (β = 0,361; EP = 0,090; p < 0,01). Isso corresponde a uma média estimada de aproximadamente 2,65 no grupo de função social, sugerindo que a evocação da missão institucional melhora de forma consistente as expectativas quanto à qualidade das análises realizadas.

Por outro lado, o enquadramento baseado em remuneração adicional não produziu efeito estatisticamente significativo sobre a qualidade esperada (β = -0,129; EP = 0,091; p = 0,154). O sinal negativo do coeficiente indica, inclusive, uma tendência – ainda que estatisticamente indistinta de zero – de redução na percepção de qualidade quando o esforço adicional é associado a incentivos financeiros, em comparação ao grupo controle.

Tomados em conjunto, os resultados revelam um padrão nítido e teoricamente consistente: incentivos extrínsecos de natureza financeira elevam a expectativa dos respondentes quanto à produtividade, mas não quanto à qualidade; já enquadramentos baseados no impacto social do trabalho elevam a expectativa quanto à qualidade, sem gerar ganhos equivalentes na produtividade esperada. Esse *trade-off* entre quantidade e qualidade é central para debates contemporâneos sobre gestão por desempenho no setor público e dialoga diretamente com críticas à adoção de bonificações financeiras como instrumento de incentivo em organizações públicas (Sousa, 2024).

Outros estudos no campo de públicas também utilizam experimentos de vinheta para examinar como pequenas variações em cenários institucionais afetam julgamentos de servidores ou gestores públicos. Migchelbrink e Van de Walle (2020), por exemplo, investigam como processos participativos afetam a disposição de oficiais públicos a utilizar contribuições da sociedade em decisões administrativas. Pedersen et al. (2018) combinam dados administrativos e experimento de *survey* para analisar como a origem étnica de usuários de serviços públicos influencia decisões discricionárias de sanção por burocratas de nível de rua em agências de emprego dinamarquesas. No contexto brasileiro, Guedes-Neto e Peters (2025) avaliam se a resistência burocrática é ativada meramente por discordâncias políticas ou se depende da percepção de ataque à democracia. Esses trabalhos ilustram a

flexibilidade do desenho de vinheta para investigar julgamentos administrativos em situações nas quais regras formais, discricionariedade, características dos usuários e formas de comunicação podem alterar a interpretação dos respondentes.

4.3 Experimentos de checagem de fatos

Os experimentos de checagem de fatos (*fact-checking experiments*) aplicados em *surveys* partem explicitamente da premissa de que indivíduos formam julgamentos e atitudes com base em crenças prévias (*priors*), que podem ser incompletas, imprecisas ou mesmo equivocadas (Nyhan & Reifler, 2010, 2015; Porter et al., 2018). Nesse tipo de desenho experimental, o grupo controle não recebe qualquer informação e, portanto, responde às perguntas mobilizando exclusivamente esses *priors*. Já o grupo tratamento é exposto a uma informação verificável e substantivamente correta – seja sobre dados, regras, políticas ou organizações –, permitindo avaliar se e como essa informação altera avaliações, percepções ou atitudes.

A lógica causal do método é direta. Na ausência de efeitos estatisticamente significativos, infere-se que os respondentes já possuíam crenças bem-informadas sobre o objeto avaliado ou que a correção não resultou em qualquer diferença no desfecho mensurado. Por outro lado, a presença de efeitos indica que a informação fornecida ao grupo tratamento corrigiu *priors* previamente imprecisos, deslocando sistematicamente as avaliações dos participantes. Diferentemente dos experimentos de vinheta ou de lista, portanto, os *fact-checking experiments* não testam preferências latentes nem julgamentos normativos sob distintos enquadramentos, mas sim processos de atualização de crenças diante de informação correta.

Em sua dissertação de mestrado, Vinicius Zimmermann (2025) aplicou um experimento de checagem de fatos com servidores da Secretaria de Estado de Fazenda do Rio de Janeiro (Sefaz-RJ). O objetivo do estudo foi avaliar se os servidores de fato conheciam a missão institucional do órgão e, em caso negativo, como a exposição ao texto formal da missão afetaria suas percepções organizacionais. A premissa substantiva do experimento é de que, em muitos contextos organizacionais, crenças sobre missão e desempenho institucional são construídas informalmente a partir da experiência cotidiana de trabalho, das

interações internas e de percepções gerais sobre a organização – e não do conhecimento explícito de documentos formais.

No experimento, os respondentes foram aleatoriamente alocados em dois grupos. O grupo controle não recebeu qualquer informação antes de responder às perguntas, sendo levado a mobilizar exclusivamente suas crenças prévias sobre a missão da organização, isto é, a partir da solicitação de pensar na missão organizacional da Sefaz-RJ sem indicar objetivamente qual era esta. Já o grupo tratamento foi exposto a um parágrafo contendo o texto oficial da missão, conforme documento institucional vigente à época da pesquisa:

> Orientar, acompanhar e avaliar o registro dos atos e fatos da Gestão Orçamentária, Financeira e Patrimonial dos Órgãos e Entidades da Administração Pública Estadual, de forma a produzir informações para tomada de decisão pelos Gestores, para o Controle Interno, Externo e Social.

Após a leitura do texto (no caso do grupo tratamento), ambos os grupos avaliaram duas afirmações em escalas numéricas de 0 a 10. A primeira mensurava o alinhamento subjetivo individual à missão institucional (“Sinto-me alinhado à missão da Sefaz”). A segunda captava a percepção sobre a capacidade organizacional (“A Sefaz consegue alcançar a sua missão”). Essas duas variáveis constituem os desfechos (*outcomes*) centrais do experimento.

Do ponto de vista metodológico, a comparação entre os grupos permite estimar diretamente o efeito causal da informação corretiva, ou seja, do conhecimento explícito da missão organizacional, sobre as avaliações dos servidores. Caso os respondentes já possuíssem conhecimento preciso sobre a missão da Sefaz, a exposição ao texto não deveria produzir qualquer efeito mensurável.

Os resultados indicam efeitos negativos da exposição à missão institucional sobre ambos os desfechos analisados. No caso do alinhamento individual à missão, o intercepto (α) estima uma média de 7,52 pontos no grupo controle (EP = 0,26), enquanto o coeficiente do tratamento é negativo e estatisticamente significativo (β = -0,77; EP = 0,38; p = 0,043), implicando uma média estimada de aproximadamente 6,75 pontos no grupo tratado ($\alpha + \beta$).

De forma semelhante, para a percepção de que a Sefaz consegue alcançar sua missão, o grupo controle apresenta média estimada de 6,50 pontos ($\alpha = 6{,}50$; EP = 0,26), ao passo que a exposição ao texto da missão reduz essa avaliação em cerca de 0,72 ponto ($\beta = -0{,}72$; EP = 0,38), resultando em uma média de aproximadamente 5,78 pontos no grupo tratamento, com significância estatística marginal ($p = 0{,}059$).

Esses achados sugerem que a comunicação explícita da missão organizacional não reforçou o alinhamento ou a confiança institucional dos servidores, mas produziu o efeito oposto. A interpretação substantiva mais plausível, conforme discutido por Zimmermann (2025), é de que os servidores possuíam *priors* relativamente positivos – ainda que imprecisos – sobre a missão da Sefaz. Ao serem confrontados com o conteúdo formal da missão, esses *priors* foram corrigidos, levando a avaliações mais críticas tanto do próprio alinhamento quanto da capacidade organizacional percebida.

Do ponto de vista analítico, o experimento ilustra de forma objetiva o potencial dos *fact-checking experiments* para identificar discrepâncias entre crenças subjetivas e informações institucionais formais no setor público. Mais do que testar preferências normativas ou respostas a incentivos, esse tipo de desenho permite examinar como a informação afeta identidades organizacionais e percepções de capacidade estatal. O fato de a Sefaz-RJ ter reformulado sua missão institucional posteriormente – ainda que sem relação causal direta com a pesquisa – reforça a relevância empírica do experimento e a importância de tratar a comunicação organizacional como um objeto de investigação causal, e não como um pressuposto dado.

Embora ainda menos frequentes na administração pública, experimentos de checagem de fatos possuem ampla aplicação em estudos sobre opinião pública, comunicação política e políticas públicas. Nyhan e Reifler (2010) oferecem um exemplo clássico ao testar se correções inseridas em notícias simuladas reduzem percepções políticas incorretas, mostrando que correções nem sempre eliminam crenças falsas e podem, em alguns casos, gerar resistência. Em estudo aplicado à política de saúde, Nyhan et al. (2013) testam se correções sobre a reforma do sistema de saúde norte-americano reduzem a crença no mito dos *death panels*. Mais recentemente, Aruguete et al. (2025) analisam como mensagens de *fact-checking* na Argentina afetam a reputação percebida de organizações checadoras e sua localização ideológica percebida. Esses exemplos mostram que esse tipo de desenho é

especialmente útil para investigar processos de atualização de crenças diante de informação correta, inclusive em temas politicamente polarizados ou sensíveis.

5. CONCLUSÃO

Este artigo teve como objetivo apresentar um guia prático para o uso de experimentos de *survey* na pesquisa em administração pública, com foco especial no contexto brasileiro. Ao longo do texto, argumentei que experimentos constituem uma estratégia metodológica particularmente adequada para investigar fenômenos centrais do campo – como percepções, julgamentos e comportamentos de servidores públicos – ao permitir inferências causais com custos relativamente baixos e elevado controle analítico. Diferentemente de abordagens exclusivamente observacionais, este método possibilita isolar mecanismos específicos, testar hipóteses substantivas de forma transparente e produzir evidências diretamente relevantes para o desenho e a avaliação de políticas públicas e organizacionais.

A partir da apresentação de três tipos de experimentos amplamente utilizados na literatura – vinheta, lista e checagem de fatos –, o artigo demonstrou como essas ferramentas podem ser operacionalizadas na prática, desde a formulação dos tratamentos até a estimação e interpretação dos efeitos. Os exemplos empíricos discutidos, todos aplicados no Brasil e com servidores públicos reais, ilustram tanto o potencial analítico desses desenhos quanto seus limites substantivos.

Experimentos de lista permitiram mensurar atitudes sensíveis relacionadas à resistência burocrática; experimentos de vinheta mostraram-se úteis para identificar tensões entre produtividade e qualidade no serviço público; e experimentos de checagem de fatos evidenciaram processos de atualização de crenças organizacionais diante de informação institucional. Em conjunto, esses casos reforçam que experimentos de *survey* não apenas funcionam no contexto brasileiro, mas produzem resultados substantivamente informativos e, por vezes, contraintuitivos.

Outros desenhos experimentais também podem ser incorporados a *surveys* e têm ganhado espaço na literatura de administração pública e áreas correlatas. Entre eles, destacam-se os *conjoint experiments*, nos quais múltiplos atributos de perfis, políticas ou cenários são manipulados simultaneamente, permitindo estimar a importância relativa de

diferentes componentes de uma decisão ou julgamento (Hainmueller et al., 2014). Esse desenho tem sido usado, por exemplo, para investigar como atributos organizacionais afetam a disposição de burocratas a resistir a pressões antidemocráticas (Silveira et al., 2026) ou como características de uma ordem administrativa influenciam a propensão de servidores a adotar práticas de *guerrilla government* (Hollibaugh et al., 2019). Embora esses desenhos não sejam tratados neste guia, eles representam uma extensão relevante para pesquisadores interessados em escolhas multidimensionais e julgamentos administrativos mais complexos.

Além da contribuição empírica, o artigo buscou enfatizar boas práticas metodológicas essenciais para a condução de experimentos de *survey* na administração pública. Questões como desenho do tratamento, aleatorização, escolha adequada dos desfechos, estimação de efeitos e interpretação substantiva dos resultados foram tratadas de forma integrada, sempre com atenção aos desafios específicos de pesquisas com servidores públicos. Aspectos éticos, limitações de validade externa e potenciais resistências institucionais também foram discutidos, reforçando que o uso de experimentos não elimina a necessidade de julgamento substantivo cuidadoso, mas exige ainda mais transparência e rigor analítico.

Do ponto de vista do campo, o artigo defende que a ampliação do uso de experimentos de *survey* pode contribuir para o avanço da administração pública brasileira, tanto na pesquisa acadêmica quanto na prática da gestão pública. Para pesquisadores, trata-se de uma oportunidade de produzir inferências causais mais robustas sobre temas tradicionalmente estudados de forma descritiva. Para gestores públicos, experimentos de *survey* oferecem uma ferramenta relativamente acessível para testar intervenções, mensagens institucionais ou mudanças organizacionais antes de sua implementação em larga escala, fortalecendo uma cultura de políticas baseadas em evidências.

Por se tratar de um artigo introdutório, este texto não substitui manuais especializados de desenho experimental e de métodos experimentais em *surveys*. Pesquisadores interessados em aprofundar o planejamento, a implementação e a análise desses desenhos podem recorrer a obras mais abrangentes sobre experimentos em ciências sociais, como Morton e Williams (2010) ou Mutz (2011), além de textos específicos sobre experimentos de *survey*, lista, vinhetas e checagem de fatos citados ao longo do artigo. O objetivo deste artigo é oferecer uma porta de entrada aplicada à administração pública brasileira, indicando os principais

cuidados conceituais e operacionais para que o leitor possa avançar posteriormente em materiais mais especializados.

Por fim, o artigo sugere que o futuro dos experimentos de *survey* na administração pública passa pela integração com outras abordagens metodológicas. A combinação de experimentos com entrevistas qualitativas, estudos de caso e análises de redes pode ampliar a validade externa dos achados e aprofundar a compreensão dos mecanismos que sustentam os efeitos observados. Ao promover essa agenda metodológica plural, mas ancorada em inferência causal rigorosa, espera-se contribuir para o fortalecimento de um campo de pesquisa mais cumulativo, empiricamente informado e diretamente conectado aos desafios concretos do setor público brasileiro.

REFERÊNCIAS

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. (Orgs.). (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 (Orgs.), *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

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 (Orgs.), *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 (Orgs.), *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 (Orgs.), *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 (Orgs.), *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 (Orgs.), *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 (Orgs.), *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 (Orgs.), *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 (Orgs.), *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 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
ORCID: https://orcid.org/0000-0001-6881-1166
Doutor e mestre em Ciência Política pela University of Pittsburgh, mestre em Economia, Política e Direito pela Leuphana Universität Lüneburg, mestre em Gestão Pública e Sociedade pela Universidade Federal de Alfenas, e graduado em Ciências Econômicas pela Universidade Federal de São João del-Rei. É professor assistente de administração pública

da Escola Brasileira de Administração Pública e de Empresas da Fundação Getulio Vargas (FGV EBAPE). É co-organizador de *Inside Brazilian Bureaucracy: Politics and Policy Implementation in Brazilian States* (Routledge) e co-autor de *Bureaucratic Resistance in Times of Democratic Backsliding* (Cambridge University Press). E-mail: joao.neto@fgv.br.

DECLARAÇÃO DE CONFLITO DE INTERESSE

João V. Guedes-Neto: Conceituação (Liderança); Curadoria de dados (Liderança); Análise formal (Liderança); Investigação (Liderança); Metodologia (Liderança); Administração de projeto (Liderança); Recursos (Liderança); Software (Liderança); Supervisão (Liderança); Validação (Liderança); Escrita - rascunho original (Liderança); Escrita - revisão e edição (Liderança).

DECLARAÇÃO DE CONFLITO DE INTERESSE

O autor não declara nenhum conflito de interesse.

DECLARAÇÃO DE DISPONIBILIDADE DE DADOS DA PESQUISA

Todo o conjunto de dados que dá suporte aos resultados deste estudo está disponível mediante solicitação ao autor. O conjunto de dados não está publicamente disponível devido a inclusão de dados secundários disponibilizados para o autor para fins deste artigo.

DECLARAÇÃO DE USO DE IA

A ferramenta de inteligência artificial ChatGPT foi utilizada para auxiliar na revisão gramatical do texto.

Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores declaram que uma vez que um manuscrito é postado no servidor SciELO Preprints, o mesmo só poderá ser retirado mediante pedido à Secretaria Editorial do SciELO Preprints, que afixará um aviso de retratação no seu lugar.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença Creative Commons CC-BY.
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.